International Journal of Bifurcation and Chaos, Vol. 21, No. 12 (2011) 3389–3415

DOI: 10.1142/S0218127411030635

CHAOS AND WEATHER FORECASTING: THE ROLE OF THE UNSTABLE SUBSPACE IN PREDICTABILITY AND STATE ESTIMATION PROBLEMS

ANNA TREVISAN
Institute of Atmospheric Science and Climate ISAC-CNR, via P. Gobetti 101, 40129 Bologna, Italy
a.trevisan@isac.cnr.it

LUIGI PALATELLA
CNISM – Dipartimento di Ingegneria dell'Innovazione, Università del Salento, Str. Prov. Lecce-Monteroni km 1,200 73100 Lecce, Italy
luigi.palatella@yahoo.it

Received October 7, 2010; Revised March 16, 2011

In the first part of this paper, we review some important results on atmospheric predictability, from the pioneering work of Lorenz to recent results with operational forecasting models. Particular relevance is given to the connection between atmospheric predictability and the theory of Lyapunov exponents and vectors. In the second part, we briefly review the foundations of data assimilation methods and then we discuss recent results regarding the application of the tools typical of chaotic systems theory described in the first part to well established data assimilation algorithms, the Extended Kalman Filter (EKF) and Four Dimensional Variational Assimilation (4DVar). In particular, the Assimilation in the Unstable Space (AUS), specifically developed for application to chaotic systems, is described in detail.

Keywords: Predictability; data assimilation; Lyapunov vectors.

1. Introduction

The belief that the atmosphere is chaotic, as Lorenz first conjectured in 1963 [Lorenz, 1963], has been consolidated throughout decades of forecasting practice. Numerical weather prediction (NWP) models are basically a nonlinear forced and dissipative system of equations, with a huge number of degrees of freedom, governing the evolution of the atmospheric variables on a discrete domain. As a consequence of the nonlinearity and instability properties of the equations, atmospheric models are subject to sensitivity to initial condition errors. The models themselves have errors associated, for instance, with truncation and the uncertainty in the parameterization of unrepresented physical processes. But, unless we wish to maintain that the state-of-the-art model at the European Centre, and competitive models at the National Meteorological Center in Washington and other centers, do not really behave like the atmosphere, in spite of the rather good forecasts that they produce at short range, we are more or less forced to conclude that the atmosphere itself is chaotic [Lorenz, 1993].

It is worth noting that the real atmosphere has more instabilities than NWP models and in this sense, due to the highly nonlinear and unstable

character of motions on small, unresolved and unrepresented scales it can be said to be "more chaotic" than NWP models.

The estimate of the initial condition, the current state, is inevitably inexact. The observing network (surface observations, upper-air soundings by radiosondes and aircrafts, satellite-based observations) is incomplete and affected by errors. Data assimilation methods combine observations with the system's evolution equations (the NWP forecast model) to obtain the best possible estimate of the current state, referred to as the *analysis*, used for diagnostic purposes as well as to define the initial condition for the forecast [Talagrand, 1997].

Originating from the initial condition and the use of imperfect models, forecast errors grow with increasing forecast lead time due to chaos. Integrations performed with the same NWP model (*perfect model* experiments) and starting from slightly different initial states reveal the following: in a finite time, the initial states develop into states that are no more similar than randomly chosen states; their distance approaches the size of the attractor. With initial condition perturbations typical of the (current) analysis error, the departures of trajectories of the same model (as well as the forecast error) approach the saturation value in a couple of weeks. The perturbations' growth depends on the particular state of the attractor of the system; forecasters have been long aware of this *flow-dependent* character of forecast uncertainty and *prediction of predictability* [Kalnay & Dalcher, 1987; Lorenz, 1996] has become an important issue in NWP.

The dynamics of infinitesimal uncertainties about a state on the attractor is governed by the tangent linear propagator. In weather forecasting context, analysis errors only in the large, meteorologically relevant *synoptic* scales are sufficiently small that their evolution is approximately governed by linearized equations for the first forecast days. The early growth of forecast errors, i.e. their growth rate and their structure, depends both on the initial state and on the structure of the initial error, on whether, for example, it assumes its greatest amplitude in synoptically active or inactive regions. In principle, one could predict the growth of infinitesimal perturbations if their projection on the unstable, neutral and stable directions was known.

In forecasting practice, the problem of predicting uncertainty has been tackled by *ensemble prediction*. In ensemble prediction, the forecast of an event is accompanied by an estimate of the probability density of its occurrence. The prediction of forecast uncertainty can be investigated theoretically by means of the Fokker–Planck equation. In NWP, there are at least two types of challenges that ensemble prediction has to face: the number of degrees of freedom of the models is huge and only approximate estimates of the amplitude and structure of the analysis error can be obtained. Among the infinite number of states that resemble the analyzed state a few have to be chosen and, in what is referred to as "Monte Carlo forecasting", perturbations are chosen at random. In a different approach, that has become the current practice, the states are chosen by adding to the analysis *dynamically constrained* perturbations, i.e. perturbations in the most unstable directions of the system. For a review of *Ensemble Prediction Systems* and ensemble initial condition generation methods, see [Leutbecher & Palmer, 2008] and [Wei *et al.*, 2008].

The role of instabilities in weather prediction, from predictability studies to ensemble forecasting and data assimilation is ubiquitous and their description in terms of Lyapunov vectors will be the *leitmotiv* of this presentation. A brief review of atmospheric predictability, from studies with simple dynamical systems to results of applications with operational NWP models will be presented in Sec. 2. Some current challenges in research on data assimilation aimed to account for the existence of an attractor and an unstable subspace will be presented in Sec. 3. In Sec. 4, we briefly deal with the connection and the interaction between chaotic properties and model error of NWP models. In Sec. 5, the conclusions will be presented together with some indications of future developments.

2. Atmospheric Predictability

The following summary is largely based on the work of Lorenz, whose contribution in the field, starting with his seminal work of 1963 [Lorenz, 1963], guided all the steps that led to our present knowledge on atmospheric predictability. In his famous talk, presented in 1972 at the 139th meeting of AAAS: "Predictability: Does the Flap of a Butterfly's Wings in Brazil Set Off a Tornado in Texas?" Lorenz discussed the following issue. "The question that really interests us is whether two particular weather situations differing as little as the immediate influence of a single butterfly will generally after sufficient time evolve into two situations differing as much as the presence of a tornado. In more

technical language, is the behavior of the atmosphere unstable with respect to perturbations of small amplitude? Although the bulk of our conclusions are based upon computer simulations of the atmosphere, the evidence that it is so is overwhelming." When predictability experiments had been conducted with the then existing general circulation models (GCMs), it was concluded that the *doubling time* of small errors was of the order of three days [Smagorinsky, 1963]. Since then, numerical weather prediction has made enormous progress, but the *Butterfly effect* is still alive and well. If a single variable (e.g. temperature) at a single grid-point, out of the 10^8 grid-points of a global NWP model (ISAC-GLOBO model) is perturbed by as little as its typical observation error (1°C), (which is much more than the effect of a butterfly's flap, but enormously less than the global uncertainty in the estimation of the state), in a matter of days appreciable differences will appear in the weather patterns throughout large portions of the globe.

2.1. *Quantifying atmospheric predictability*

Most of the attempts to quantify atmospheric predictability are based on models. The more realistic the models have become, in the sense that they produce better and better forecasts, the more confident we are that their properties are similar to those of the real system.

Parallel to the improvements in the skill of weather forecasts, due to the reduction of model and initial condition errors, the theory of dynamical systems and predictability has made substantial progress. The early attempts to quantify atmospheric predictability were concerned with the average growth of errors with forecast lead time. The relevance of Lyapunov exponents and vectors in the problem was not clear or accepted at the time: the main source of skepticism was the fact that Lyapunov exponents were considered a measure of the global average long-term growth rate of infinitesimal perturbations. It took several years before the role of Lyapunov exponents and vectors in the early stages of error growth became established.

The leading Lyapunov exponent appeared as an important ingredient in Lorenz's first attempt [Lorenz, 1969] to infer atmospheric predictability from climatic data by looking at the evolution of *analogs*, observed atmospheric states that resemble one another as closely as can be found in the climatic record. Lorenz refers to this work as the *empirical approach* as opposed to the *dynamic approach* that uses models to determine the growth rate of small errors. He looked at the time evolution of the distance, the root mean square (rms) difference between states (referred to as the error) between the closest couples of states found in the record, averaged over an ensemble of time-decorrelated analogs. In this way, the rate of divergence of trajectories that are on the atmospheric attractor could be directly estimated. However, given the short available climatic record, finding *good* analogs (i.e. couples of states whose distance from each other is small) is essentially impossible, so that the growth rate of small errors could only be inferred from the growth rate of large errors. To this end, Lorenz assumed that nonlinearity associated with advection introduces a quadratic term in the equation describing average error growth; in this equation, the time derivative of the average error E is the sum of a linear term with a coefficient given by the leading Lyapunov exponent and a quadratic term that describes nonlinear saturation:

$$\frac{dE}{dt} = \lambda_1 \left(\frac{E - E^2}{E_\infty} \right). \tag{1}$$

With this empirical approach, by fitting the parameters of the error growth curve to the data, Lorenz obtained an estimate of the leading Lyapunov exponent corresponding to a doubling time of 2.5 days. It should be mentioned that the resulting exponent is the one associated with atmospheric instabilities present on the large scales that could be resolved by the data and by the models that were run at the time. The equation, later referred to as *Lorenz's law* of error growth has been used by several authors over time, up to the present, to describe the behavior of errors from the early stages of their growth up to the forecast time when they reach saturation and cease to grow.

Lorenz's law has been extensively used to describe the error growth within NWP models (with a resolution that increased throughout the years of forecasting practice), by applying it to average rms forecast differences. It has also been used, with some modifications, to describe the growth of forecast error, by applying it to average rms differences between forecast and analysis states. For example, Kalnay and Dalcher [1987] introduced an additional constant term in order to account for the effect of model deficiencies on the growth of forecast error.

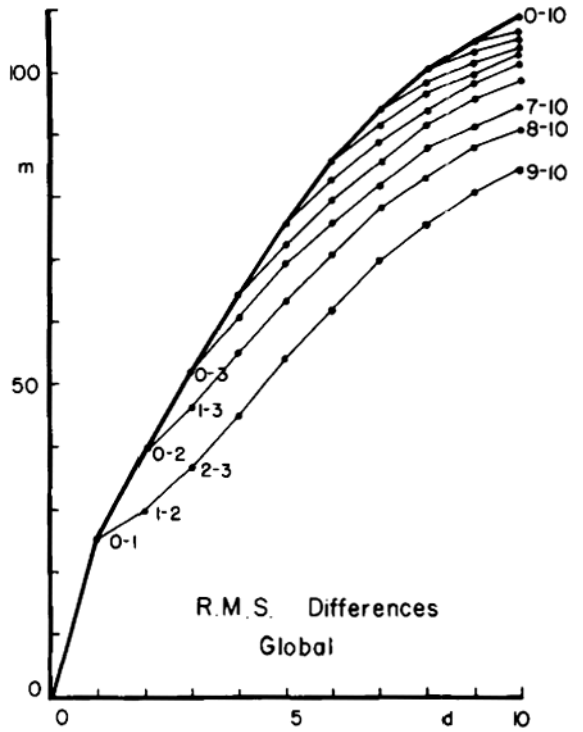

Fig. 1. Global root-mean-square 500 mb height differences E_{jk} in meters, between j-day and k-day forecasts made by the ECMWF operational model for the same day, for $j < k$, plotted against k. Values of (j, k) are shown beside some of the points. Heavy curve connects values of E_{0k}. Thin curves connect values of E_{jk} for constant $k-j$. After [Lorenz, 1982].

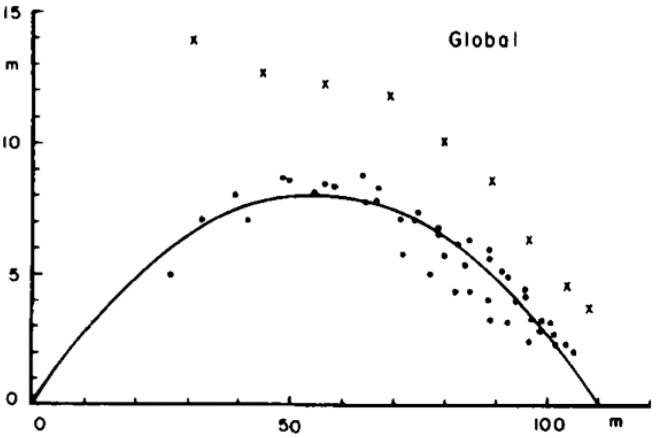

Fig. 2. Increases in global root-mean-square 500 mb height differences, $E_{j+1,k+1} - E_{jk}$ plotted against average height differences $(E_{j+1,k+1} - E_{jk})/2$, in meters, for each one-day segment of each thin curve in Fig. 1 (large dots), and increases in $E_{0,k+1} - E_{0,k}$ plotted against average differences $(E_{0,k+1} + E_{0,k})/2$ for each one-day segment of heavy curve in Fig. 1 (crosses). Parabola of best fit to large dots is shown; after [Lorenz, 1982].

Figures 1 and 2 present the results obtained in [Lorenz, 1982] using the ECMWF model forecasts. In Fig. 1 the error growth is shown for actual forecast errors (as difference between forecast and analysis states) and for a perfect model scenario (as difference between "lagged forecasts states" verifying on the same day). Figure 2 shows the fit of Eq. (1) to the perfect model data of Fig. 1.

2.2. *Early stages of error growth*

The inspection of the early stages of error growth revealed that the problem is more complex and that the average growth rate of infinitesimal uncertainties cannot be described simply by the leading Lyapunov exponent: the average growth rate has a time dependent behavior that depends upon the structure of the initial perturbation.

One of the first attempts to estimate the time dependence of the growth rate of small random perturbations (obtained by rescaling the difference from a time-decorrelated state on the attractor of the model), in a realistic general circulation model is found in [Schubert & Suarez, 1989]. They observed that the average (over an ensemble of initial states) growth rate $(1/E)(dE/dt)$ increased with time during the first few days and then started to decline when the error was as small as 1% of its saturation value, presumably still in a regime of growth governed by linearized dynamics. Clearly the initial growth of small perturbations in the model (Fig. 3) is not described by the linear term in Eq. (1).

Several authors [Mukougawa *et al.*, 1991; Trevisan *et al.*, 1992; Nicolis & Nicolis, 1993; Trevisan, 1993] conducted systematic studies in low order dynamical systems, where the leading Lyapunov exponent (LLE), λ_1 could be independently estimated. The aim was to investigate how the phase space average growth rate of small perturbations

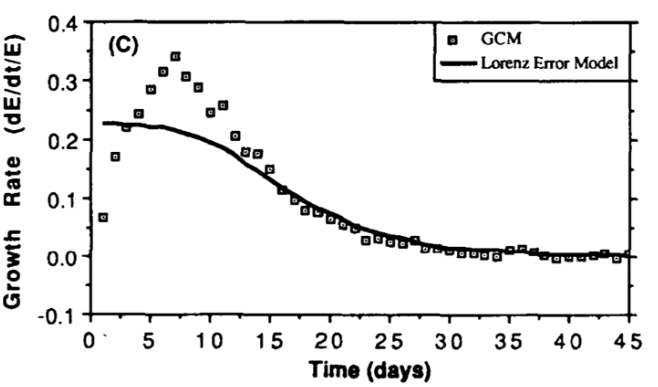

Fig. 3. After [Schubert & Suarez, 1989]. The growth rate of the error $(1/E)(dE/dt)$ as a function of forecast time with the best fit to the Lorenz error model.

(obtained by averaging over a very large number of initial conditions) compared with λ_1. Departures from Lorenz's law in the initial stages of error growth were observed; error dynamics was found to display a transient regime of growth, with growth rates that may exceed the LLE, before settling in the mature Lyapunov stage, with constant growth rate λ_1.

Local and global predictability properties were studied in [Trevisan & Legnani, 1995] by making a distinction between two types of variability: phase space variability, associated with the local instability properties (local Lyapunov exponents themselves vary in phase space) and the variability associated with the direction of the initial perturbation. Most importantly, the transient behavior of phase space average error growth rate was found to depend upon the structure of initial perturbation; moreover, the average growth rate can initially be larger than the LLE. These results were confirmed in the study of a high-dimensional atmospheric model [Vannitsem & Nicolis, 1997], where the role of the dominant (baroclinic) instability on the spectral distribution of Lyapunov vectors was investigated. Szunyogh *et al.* [1997] compared the growth and structure of random, singular vectors and Lyapunov vectors perturbations in a low-resolution GCM (these types of perturbations were used at the time and are still used in the operational practice for initializing members for ensemble prediction). They discussed the important role of Lyapunov vectors in linearized perturbation development in the context of ensemble prediction.

The time evolution of average prediction errors will be discussed in the following sections together with some background on Lyapunov and Singular vectors. Important aspects of error growth, related to its variability around the mean (see e.g. [Nicolis, 1992]) and the relationship between ensemble spread and forecast error (see e.g. [Whitaker & Loughe, 1998]) will not be covered here. Most of the results presented come from early works with simple systems and refer to average error growth of random and singular vector perturbations compared with the LLE. To the authors' knowledge, analogous results with NWP models are not available; given the need to visit the whole attractor and the slow convergence of numerical algorithms experienced even in models with a small number of degrees of freedom, an estimate of the LLE of a NWP model would probably be computationally very demanding.

2.2.1. *The growth of random perturbations*

Figure 4 shows the error growth rate as a function of time in [Lorenz, 1963] model for three different initial error amplitudes. Starting with initial errors in a random direction, due to the existence of stable directions, depending on the system, the average growth rate is very small or even negative initially. A second stage of transient *super-Lyapunov* growth is observed, during which the average growth rate is larger than the LLE; this stage lasts until perturbations align along the direction of the leading Lyapunov vector; in the third stage, the average growth rate is approximately equal to the LLE. These are the three stages that can be observed in simple systems when error dynamics is linear.

When nonlinearity sets in, at a time that depends upon the initial amplitude of the error, the growth starts to decline and ceases completely when the error reaches saturation; if the error is not sufficiently small initially, one cannot observe all the stages of growth of the linearized dynamics. The growth rate curve corresponding to the largest amplitude (0.1) in Fig. 4 first increases and then starts to decrease before settling in the mature Lyapunov stage so that only the first two stages of growth are present. The growth rate curve of Fig. 3

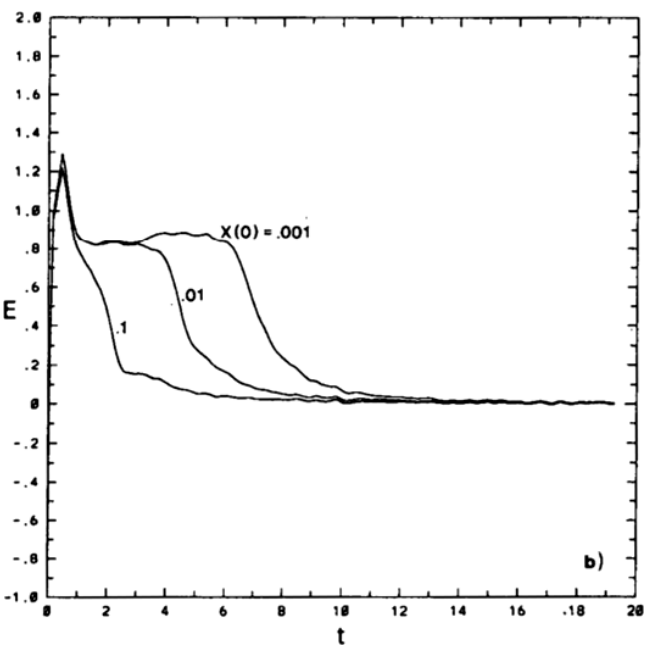

Fig. 4. After [Trevisan, 1993]. Error growth rate as a function of the time for different values of the starting amplitude $X(0)$ for the [Lorenz, 1963] model.

has a similar behavior, but the estimate of the LLE is obtained by fitting Lorenz's law and it is hard to draw any firm conclusion from the similarity of the two curves.

We have computed the growth rate from "lagged forecast" errors of Fig. 1 and from similar figures from more recent versions of the ECMWF model and observed that the growth rate monotonically decreases with time. We can conclude from [Lorenz, 2005] that with a realistically large initial uncertainty, it is quite possible that the slackening of the growth rate will commence just as the early rapid growth is ending, and the mature Lyapunov stage will be virtually absent. An alternative explanation is that we do not see the plateau of the mature Lyapunov stage because averaging is not made over a large enough sample of initial conditions.

2.2.2. *The growth of singular vector and Lyapunov vector perturbations*

Perturbations in the direction of the leading singular vectors (SVs) have a different behavior from that of random perturbations; transient growth of SV perturbations is very large at initial time and monotonically decreases towards the leading Lyapunov exponent, λ_1 (see Fig. 5 and [Trevisan & Legnani, 1995]).

Singular values and vectors of the tangent linear propagator, $\mathbf{M}$, were introduced in [Lorenz, 1965] to describe the evolution of errors towards their asymptotic directions. If an infinite collection of states fills the surface of an infinitesimal sphere in phase space, the sphere is transformed after a given time interval, into an ellipsoid, whose semiaxes are the left (final) singular vectors of $\mathbf{M}$ multiplied by the respective eigenvalues (see Sec. 2.3.2 for details). The right (initial) singular vector associated to the largest eigenvalue is the perturbation that is transformed into a perturbation in the direction of the largest axis of the ellipsoid.

After some time, the singular vectors with singular values larger than one, span the same subspace as Lyapunov vectors with positive exponents: asymptotically, all vectors align in the direction of the leading Lyapunov vector. In the presence of a single dominant instability, the time needed for convergence to the leading Lyapunov vector is quite short. In the quasi-geostrophic (QG) model studied in [Vannitsem & Nicolis, 1997], that describes baroclinic instability (dominant in the atmosphere

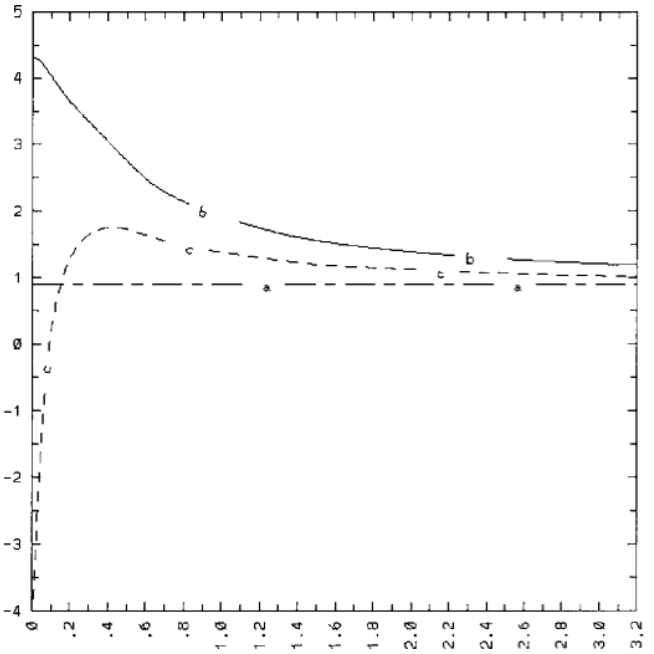

Fig. 5. (a: long-short dash) Global Lyapunov exponent, λ_1; (b: solid) ensemble average growth rate of the most unstable perturbation as a function of time and (c: short dash) ensemble average growth rate of random perturbation. Averages are over the whole attractor phase space. After [Trevisan & Pancotti, 1998].

and responsible for the development of weather systems), the time necessary for all perturbations to grow at the rate of the LLE is only about twice λ_1^{-1}.

Lacarra and Talagrand [1988] observed that, if the eigenvectors of a constant tangent linear propagator, $\mathbf{M}$ are not orthogonal, for finite times the growth of perturbations exceeds that of the leading eigenvector of $\mathbf{M}$. The directions in which an initial infinitesimal error is most rapidly amplified can be determined by maximizing the amplification factor, $A(\delta\mathbf{x})$:

$$A(\delta\mathbf{x}) = \frac{\langle \mathbf{M}\delta\mathbf{x}, \mathbf{M}\delta\mathbf{x}\rangle}{\langle \delta\mathbf{x}, \delta\mathbf{x}\rangle} = \frac{\delta\mathbf{x}^T\mathbf{M}^T\mathbf{M}\delta\mathbf{x}}{\delta\mathbf{x}^T\delta\mathbf{x}}. \quad (2)$$

In order to explain super-Lyapunov growth, [Trevisan & Pancotti, 1998] extended the analysis of [Lacarra & Talagrand, 1988] to the study of the stability of a chaotic system. They discussed how the role of eigenvectors in the description of asymptotic growth for a constant $\mathbf{M}$ is played by Floquet eigenvectors and by Lyapunov vectors in the case of periodic and aperiodic trajectories, with $\mathbf{M}$ varying along the trajectory. Studying the stability of periodic and aperiodic orbits in the [Lorenz, 1963] system, they showed that transient (super-Lyapunov)

growth is a consequence of the nonorthogonality of Lyapunov vectors. They also discussed the relation between singular vectors and Lyapunov vectors (see Sec. 2.3.2).

One of the conclusions of [Trevisan & Pancotti, 1998] regards the generation of perturbations in ensemble forecasting: because of the large number of degrees of freedom of numerical prediction models, there may be a number of initial singular vectors that grow very quickly during an initial short transient but that, pointing in directions where the attractor is not continuous, are incompatible with the dynamics. The fast growth of the initial singular vectors is limited to the short time span needed for the perturbations to assume the structure of the dominant Lyapunov directions [Szunyogh *et al.*, 1997].

2.3. *Lyapunov vectors, singular vectors and their relation*

2.3.1. *Lyapunov vectors*

Lyapunov vectors, defined for nonlinear systems, are the time dependent dynamical structures associated with the Lyapunov exponents. There are basically two definitions of Lyapunov vectors; either one of these two sets of Lyapunov vectors span the same invariant Oseledec subspaces. The first definition is that of an orthonormal set of vectors, the eigenvectors of the limit self-adjoint operator $\mathbf{\Phi}_\infty(t)$ [Oseledec, 1968]:

$$\mathbf{\Phi}_\infty(t) = \lim_{t_0 \to -\infty} [\mathbf{M}_{t_0 \to t} \mathbf{M}^*_{t_0 \to t}]^{\frac{1}{2(t-t_0)}} \tag{3}$$

where $\mathbf{M}_{t_0 \to t}$ is the tangent linear operator of the temporal evolution from time t_0 to time t, while $\mathbf{M}^*_{t_0 \to t}$ is its adjoint operator with respect to the standard scalar product. The initial state $\mathbf{x}_0$ of the nonlinear trajectory is on the attractor. The Oseledec theorem proves that the limit does not depend on the choice of $\mathbf{x}_0$ except for a null measure set. For more information see [Lorenz, 1984] and [Legras & Vautard, 1996].

The second is a set of nonorthogonal Lyapunov vectors, which are independent of the norm and each of them is mapped into itself by the tangent linear propagator $\mathbf{M}_{t_0 \to t}$ [Eckmann & Ruelle, 1985; Brown *et al.*, 1991]. These vectors have been shown to be the natural generalization of eigenvectors and Floquet vectors to aperiodic flow [Trevisan & Pancotti, 1998]. An important remark is that the nonorthogonal Lyapunov vectors become the orthogonal eigenvectors of $\mathbf{\Phi}_\infty(t)$ if orthonormalization going from the largest to the smallest Lyapunov exponent is performed.

The following standard technique is commonly used for calculating the orthonormal set of Lyapunov vectors [Benettin *et al.*, 1980]: a set of N initially random tangent vectors are linearly evolved and orthonormalized every τ time units. After a spin-up time, these vectors span the N-dimensional most unstable subspace of the system.

An efficient method for recovering norm-independent nonorthogonal Lyapunov vectors is given in [Wolfe & Samelson, 2007]. Either one of the two above-mentioned methods can be used to identify the unstable, neutral, stable subspaces, the span of Lyapunov vectors with positive, null, negative exponents.

In weather prediction, bred vectors [Toth & Kalnay, 1993, 1997] are usually computed instead of Lyapunov vectors. Bred vectors are the finite amplitude generalization of Lyapunov vectors and are computed as differences between twin nonlinear model integrations. The renormalization amplitude ξ and breeding time τ are the parameters being tuned to select the instability scale. With an infinitesimal renormalization amplitude and periodic orthonormalization, the bred vector algorithm would produce the same results as the Lyapunov vector algorithm in [Benettin *et al.*, 1980].

2.3.2. *Singular vectors and nonorthogonal Lyapunov vectors*

For a definition of singular vectors we follow the work of [Lorenz, 1965]. In this paper the author studies the evolution of an infinitesimal sphere obtained by adding to the state of the n-dimensional system an error $\mathbf{y}(t_0)$ satisfying the condition

$$\mathbf{y}(t_0)^T \mathbf{y}(t_0) = \epsilon^2. \tag{4}$$

If the system equations are given by $\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x})$, the growth of the infinitesimal error $\mathbf{y}(t)$ will follow the equation

$$\mathbf{y}(t_1) = \mathbf{M}_{t_0 \to t_1} \mathbf{y}(t_0) \tag{5}$$

where the tangent linear operator $\mathbf{M}_{t_0 \to t_1}$ satisfies

$$\mathbf{M}_{t_0 \to t} = \exp\left(\int_{t_0}^{t} \frac{\partial \mathbf{F}(\mathbf{x}(\tau))}{\partial \mathbf{x}} d\tau \right). \tag{6}$$

If errors evolve according to (5) the sphere is deformed into an ellipsoid at time t:

$$\mathbf{y}(t)^T(\mathbf{M}_{t_0\to t}\mathbf{M}_{t_0\to t}^T)^{-1}\mathbf{y}(t) = \epsilon^2. \tag{7}$$

The matrix $\Phi_{t_0}(t) \equiv \mathbf{M}_{t_0\to t}\mathbf{M}_{t_0\to t}^T$ is symmetric and has positive eigenvalues denoted by Γ_i^2. Consequently, the length of the ellipsoid semi-axes will be equal to $\Gamma_i\epsilon$. The eigenvectors of $\Phi_{t_0}(t)$ identify the directions of the semi-axis and are called left (final) singular vectors with associated exponents given by $\gamma_i = 1/(t-t_0)\log(\Gamma_i)$. On the contrary, the initial (at time t_0) directions of errors that will grow at rate γ_i and that evolve into the ellipsoid semi-axis directions are the eigenvectors of $\mathbf{M}_{t_0\to t}^T\mathbf{M}_{t_0\to t}$ and are referred to as right (initial) singular vectors. This can be understood if we write the singular vector decomposition of the matrix $\mathbf{M}_{t_0\to t}$ as

$$\mathbf{M}_{t_0\to t} = \mathbf{U}\boldsymbol{\Gamma}\mathbf{V}^T \tag{8}$$

where $\mathbf{V}, \mathbf{U}$ are orthogonal while $\boldsymbol{\Gamma}$ is diagonal. Indeed we have, due to orthogonality of $\mathbf{V}, \mathbf{U}$ that

$$\begin{aligned}\mathbf{M}_{t_0\to t}\mathbf{M}_{t_0\to t}^T &= \mathbf{U}\boldsymbol{\Gamma}^2\mathbf{U}^T,\\ \mathbf{M}_{t_0\to t}^T\mathbf{M}_{t_0\to t} &= \mathbf{V}\boldsymbol{\Gamma}^2\mathbf{V}^T\end{aligned} \tag{9}$$

and it easy to see that in the first (second) case the eigenvectors are the columns of $\mathbf{U}(\mathbf{V})$. It is worth noting that for $t_0 \to -\infty$ the first N left singular vectors tend to span the same subspace as the first N Lyapunov vectors at time t. Fixing t_0 and letting $t \to \infty$ the last $n-N+1$ (where n is the total dimension of the phase-space) right singular vectors tend to span the same subspace of the last $n-N+1$ Lyapunov vectors at time t_0. Using this property Trevisan and Pancotti [1998] numerically determined the nonorthogonal kth Lyapunov vectors intersecting for the same time t the span of the first N final singular vectors with the span of the last $n-N+1$ initial singular vectors (see for example Fig. 6).

The (nonorthogonal) Lyapunov vectors referred to also as *backward* [Legras & Vautard, 1996] or *characteristic* vectors are invariant under the flow, that is, they map into themselves under the tangent linear propagator:

$$\mathbf{e}^{(i)}(t_2) = \mathbf{M}_{t_1\to t_2}\mathbf{e}^{(i)}(t_1) = \Lambda^{(i)}\mathbf{e}^{(i)}(t_2) \quad \text{with}$$

$$\Lambda^{(i)} \equiv \exp\left(\int_{t_1}^{t_2} \mathrm{d}t\lambda^i(t)\right) \tag{10}$$

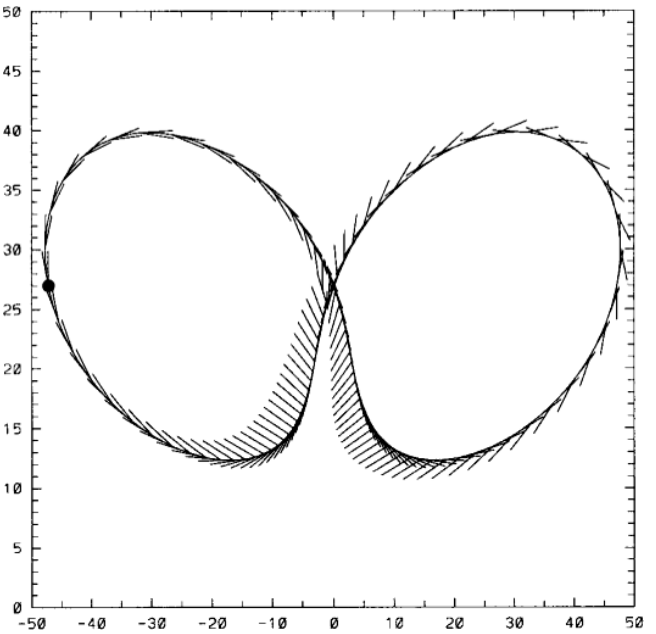

Fig. 6. After [Trevisan & Pancotti, 1998]. Periodic orbit $\bar{j}$ and first and second Lyapunov vectors, projected on the plane $x = y$. The second Lyapunov vector, corresponding to the exponent $\lambda_2 = 0$, is tangent to the flow. The orbit's period is $T = 1.558$.

where $\mathbf{e}^{(i)}(t)$ are the normalized ith Lyapunov vectors at time t and

$$\lambda^{(i)}(t) = \frac{d}{dt}\log\|\delta\mathbf{x}^i(t)\| \tag{11}$$

is the ith local Lyapunov exponent; $\delta\mathbf{x}^i(t) = \epsilon\mathbf{e}^{(i)}(t)$, with $\epsilon \ll 1$.

It needs to be stressed here that once the Lyapunov vectors have been numerically computed for a given $\mathbf{x}(t)$, it is not possible to obtain the invariance property by direct forward numerical integration. This is due to the fact that also a very small numerical error will be amplified by the tangent linear operator with consequent alignment of all the Lyapunov vectors along the leading one. For this reason, one should perform the calculation of [Trevisan & Pancotti, 1998] for each time step of the trajectory.

2.4. *The role of scales in predictability*

In simple systems or models of intermediate complexity discussed in Sec. 2, the description of the early stages of error growth and their relation to the leading Lyapunov exponent is fairly accurate. The picture is more complex in realistic contexts or models that include more than one dominant instability.

Errors in observing smaller-scale features, especially the more energetic ones, will grow rapidly; the error in observing the details of a thunderstorm should amplify at least as rapidly as the thunderstorm itself, doubling in half an hour or less. In short, the largest Lyapunov exponents of high-resolution atmospheric models, and of the atmosphere itself, are associated with small scales. These same errors, however, soon acquire limiting amplitudes, at the time when errors in the larger scales are just beginning to reveal their growth. The latter errors, aside from growing more slowly and therefore being associated with smaller Lyapunov exponents, continue to grow much longer, generally doubling in two days or more, and they ultimately acquire much larger amplitudes than small-scale errors; they therefore provide the major contribution to the total error in the forecast [Lorenz, 1986].

To make progress in the understanding of the problem there are again two approaches. The first is to resort to conceptual models as useful paradigms of the essential mechanisms; the other is to experiment in the "laboratory" provided by realistic NWP models.

2.4.1. *Lorenz's 1969 paper*

One of the first attempts to study the predictability of a flow with many scales of motion is due to [Lorenz, 1969]. Starting from the vorticity equation, new equations whose variables are ensemble statistics were derived to describe the evolution of the spectral density of the *error energy*, the square norm of the error in the velocity field. As a closure, quadratic functions of the errors and quadratic functions of the basic state were assumed to be independent and the spectral density of the basic state energy was specified in advance. The derived equations were formally linear and, to account for nonlinear effects, the error energy in one scale was replaced by a constant as soon as it reached the spectral energy of the basic state.

The shape of the kinetic energy spectrum exerts a controlling influence on the time scale associated with each spatial scale; each scale has an intrinsic range of predictability provided the total energy of the system does not fall off too rapidly with decreasing wavelength. Two important features of the observed atmospheric spectrum are the synoptic scale k^{-3} range of quasi-geostrophic turbulence (3000–600 km) and the $k^{-5/3}$ mesoscale range (200–2 km). Lorenz [1969] assumed a spectral shape resembling the observed spectrum, with a $-5/3$ tail and computed the growth of an error field confined initially to the smallest scales, showing how the error contaminates the whole spectrum in a matter of days. The model is crude in many respects and very dependent on the assumed shape of the spectrum but, overall, the result illustrates how the growth of errors in the larger scales is influenced by the smaller scales. Subsequent studies, using more sophisticated closure assumptions [Leith, 1971; Leith & Kraichnan, 1972] have yielded qualitatively similar results. With different shapes of the spectrum, some containing a gap in the *mesoscale*, compatible with the imperfect knowledge of the real atmospheric spectrum in that band, Lorenz found results compatible with the growth of errors in the ECMWF model and concluded that there is a limit (of the order of two weeks) to the predictability time of large scale weather patterns (the predictability time can be defined as the time after which the average distance between the forecast and the actual state of the system is equal to — or in practice a large percentage value of — the average distance between two randomly chosen states).

Lilly [1990] discussed Lorenz's results on the predictability limit from the turbulence theory perspective. In the inertial range of a turbulent fluid, scaling arguments suggest that a characteristic time scale, the "Eddy turn-over time", T_e is a function of the horizontal wave number k and kinetic energy $E(k)$:

$$T_e(k) = \frac{1}{[k^3 E(k)]^{1/2}}. \tag{12}$$

If errors at scale k contaminate errors in the next larger scale in an eddy turn-over time, the predictability time of a particular scale $T_p(k)$ i.e. the total time for errors to propagate from the smallest resolved scale k_r to the scale k is given by:

$$T_p(k) = \int_k^{k_r} \left(\frac{T_e}{k'}\right) \mathrm{d}k' = \int_k^{k_r} \frac{1}{k'^{5/2} E(k')^{1/2}} \mathrm{d}k'. \tag{13}$$

If the spectrum is proportional to k^{-n}, with $n > 3$, we obtain

$$T_p(k) \propto \int_k^{k_r} \frac{1}{k'^{(5-n)/2}} \mathrm{d}k' = \frac{n-3}{2} [k'^{(n-3)/2}]_k^{k_r} \simeq \frac{n-3}{2} k_r^{(n-3)/2}. \tag{14}$$

Thus $T_p(k)$ is proportional to $k_r^{\frac{n-3}{2}}$. Under this condition the predictability increases for all scales k when the resolution k_r is improved. If $n < 3$ however the integral becomes

$$T_p(k) \propto \int_k^{k_r} \frac{1}{k'^{(5-n)/2}} \mathrm{d}k' = \frac{n-3}{2}[k'^{(n-3)/2}]_k^{k_r} \simeq \frac{3-n}{2} k^{(n-3)/2}. \quad (15)$$

In this condition, $T_p(k)$ is proportional to $k^{-\frac{3-n}{2}}$, so that the predictability of scale k becomes independent of k_r and consequently bounded regardless of the resolution. For a -3 slope of the spectrum, the error growth is independent of the spatial scale so that the predictability time is not limited (cutting by half the initial error, the predictability time is increased by one doubling time). For shallower spectra smaller scales grow faster; because of the spreading of the error to the neighboring scales the predictability time has a finite limit.

2.4.2. *Lorenz's 1996 toy-model*

Planetary and synoptic scale motions, resolved in global NWP models, are characterized by error doubling times of the order of 1 to 2 days, while convective scales have doubling times of the order of 1 h. Predictability in the presence of slow and fast growing modes was investigated by [Lorenz, 1996] in a simple model whose variables have two distinct time scales. In this model, two coupled sets of equations describe the evolution of variables, defined on a longitudinal grid, that mimic planetary and convective motions, with their typical time and amplitude scales. Starting from a state on the attractor and adding initial random perturbations, twin nonlinear integrations were performed to estimate the growth of errors in this model.

Ensemble average growth curves of initially small perturbations show that the convective scales dominate the initial growth, double in approximately 6 h and then reach their small amplitude saturation level. Subsequent growth is typical of the large scale, with a doubling time of about 1.6 days; all growth ceases at the large scale saturation level. The convective scale doubling time is consistent with the leading Lyapunov exponent of the coupled system. The subsequent, slower quasi-exponential growth is consistent with Lyapunov exponents of planetary scale instabilities. The model captures the essential aspects of error dynamics of more realistic atmospheric models with multiple-scale instabilities (see Fig. 7).

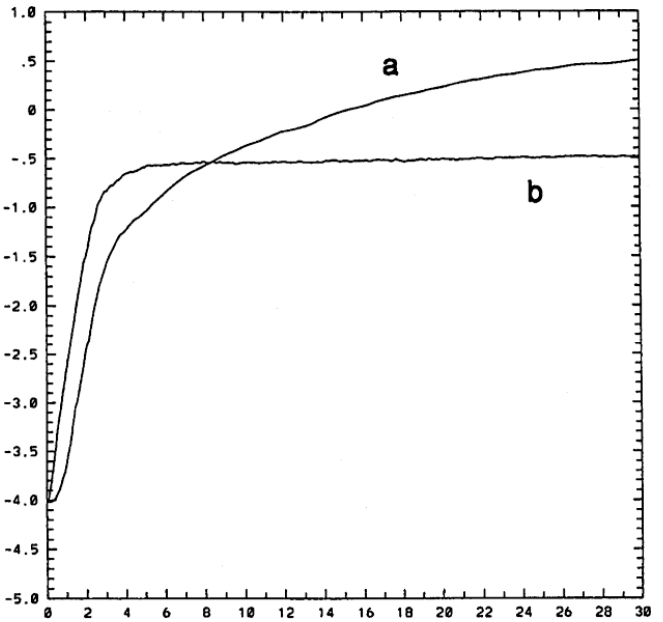

Fig. 7. After [Lorenz, 1996]. Variations of $\log_{10} E$ (scale at left) with prediction range τ (scale, in days, at bottom), shown separately for large scales (curve a) and small scales (curve b), for 30 days, as determined by 25 pairs of integrations of the model [Lorenz, 1996].

2.4.3. *The breeding method*

In realistic models of the atmosphere we expect to find a large number of positive Lyapunov exponents so that, with standard methods and a limited number of initial perturbations, only the most rapidly growing unstable vectors will emerge. In forecasting practice, before Lorenz clarified the dynamical mechanism of selection of instabilities in his 1996 model [Lorenz, 1996] researchers at the National Meteorological Center (NMC, now NCEP) were well aware of this problem and the breeding method [Toth & Kalnay, 1993, 1997] was devised for its solution.

The breeding method is the nonlinear generalization of the procedure used to compute leading Lyapunov vectors and exponents. Perturbations, computed as difference between two states belonging to nonlinear model trajectories are periodically rescaled to a fixed amplitude. Random perturbations are chosen initially, but the breeding cycle is repeated indefinitely and the perturbations gradually become structured in the most unstable

directions, within 2–3 days in global atmospheric models [Kalnay, 2003].

We have already discussed in Sec. 2.3.1 the relation between Lyapunov and breeding vectors and the role of the two breeding parameters: the rescaling amplitude ξ and the rescaling time interval; if these parameters are both sufficiently small and the perturbations are also orthonormalized, after an initial transient time interval, the first N perturbations will span the subspace of the N leading Lyapunov vectors. When there exist instabilities with different spatial and time scales, it is possible to tune the breeding parameters in order to select the slower instabilities with larger amplitude [Toth & Kalnay, 1993]. If the amplitude and rescaling interval are not small, the fast growing small scale bred modes will reach their saturation level and the larger scale instabilities will emerge, by the same mechanism as in the [Lorenz, 1996] model.

Toth and Kalnay [1993] found that when the initial amplitude was chosen to be within the range of estimated analysis errors (i.e. between 1 m and 15 m for the 500-hpa geopotential height) the bred vectors developed faster in strong baroclinic areas. Their horizontal scale was that of short baroclinic waves, and their hemispherical average growth rate was about 1.5/day (similar to the estimated growth of analysis errors). However, if the initial amplitude was chosen to be much smaller than the estimated analysis error (10 cm or less), then a different type of bred vector appeared, associated with convective instabilities, which grew much faster than baroclinic instabilities (at a rate of more than 5/day). The faster instabilities saturated at amplitudes much smaller than the analysis error range.

The nonlinear saturation of irrelevant fast growing modes is an advantage that suggests the use of breeding for other problems. For example, for seasonal and inter-annual forecasting using a coupled ocean-atmosphere system, the slower growing (but very large amplitude) coupled ENSO instabilities could be captured, while eliminating through nonlinear saturation the irrelevant details of weather perturbations [Cai *et al.*, 2003; Kalnay, 2003].

2.5. *Predictability studies from global NWP models*

New challenges in predictability research are continuously brought up by the development of assimilation and forecasting techniques. Improvements in the skill of NWP model forecasts are related to the reduction of errors in the initial condition and in the model. Improved accuracy in the initial condition is due not only to the enhancement of the observational network but also to the development of advanced data assimilation techniques. Improvements in the model are due to the inclusion of better representation of physical processes as well as to the development of ensemble forecasting techniques and increased model resolution.

Due to the increased resolution of global NWP models the resolved spatial scales have moved to higher and higher wave number ranges with consequent reduction of the initial condition error. Below we report the results of two NWP model studies of the role of scales in predictability.

2.5.1. *Band limited and small scale error growth*

Tribbia and Baumhefner [2004] investigated the growth and scale propagation of error energy initially confined in different ranges of spatial scales. They used the NCAR Community Climate Model Version 3 (CCM3), at various horizontal resolutions ranging from T42 (where "T" stands for triangular truncation) to T170 (corresponding roughly to 280–75 km) to conduct identical model and imperfect-model twin experiments.

In the perfect model twin experiments (T63), the model initial condition was perturbed by adding initial errors with amplitude and spatial structure with the same characteristics of the estimated analysis error. After approximately one day (a transient time necessary for the perturbations to acquire the structure of unstable modes) the growth of errors was approximately exponential and disturbances have a spectral peak in synoptic scales (two-dimensional wave numbers 10–20) typical of baroclinic instability. If the initial error was band limited to wave numbers $n < (> 30)$, the error propagates upscale (downscale) to end up with the same spectral distribution peaked at synoptic scales, but with different amplitude: the spectral amplitude is smaller if n was initially < 30, while for $n > 30$ the amplitude is approximately the same as if the initial error was not band limited. Therefore, very little is gained by accurately specifying the small scales. Due to the rapid transfers of errors across scales and the "seeding" of baroclinically active scales, the same authors arrived at the conclusion that errors due to truncation were not as

severe as initial condition errors for the prediction of synoptic scale weather.

The results presented in this paper can be interpreted in the light of previous dynamical systems studies. Errors that are initially random, after some transient time acquire the structure of the leading Lyapunov vectors, the unstable directions of the aperiodic trajectory. As discussed in previous sections, in global NWP models that do not explicitly describe convection the dominant instabilities are baroclinic. Therefore, whether the initial condition error is distributed in all spatial scales or is band limited to the short or the large scales, we expect that after a while the error will have the same spectral distribution of the dominant unstable structures. This is confirmed in [Tribbia & Baumhefner, 2004] study: after one forecast day, errors that were band-limited have a similar spectral distribution and grow exponentially at about the same rate; after three forecast days their spatial structure (and that of the error that was not initially band-limited) shows very similar features all over the globe (see Fig. 5 of [Tribbia & Baumhefner, 2004]). In summary, after an initial transient time of the order of one day, regardless of the initial spectral distribution of the error energy, the structure and growth of the error is that of the baroclinically active synoptic scales.

Harlim *et al.* [2005] investigated the growth of infinitesimal random errors ($10^{-3}E_\infty$) in the National Center for Environmental Prediction (NCEP) Global Forecast System (GFS). Starting with such small initial perturbation amplitudes, tropical convection that saturates at less than $10^{-3}E_\infty$ [Toth & Kalnay, 1993, 1997] gives rise to super-fast growth that lasts approximately 1 h. In a second phase, that lasts approximately 1 day, small scale turbulent dynamics dominates, doubling times increase from 3 h to approximately 1 day while the error increases to values of the order $10^{-2}E_\infty$. During this stage, the growth rate decays in a log-linear fashion, that can be approximated by $(1/E)(dE/dt) = \lambda_f \log(E/E_f)$. According to this fit, the error growth would cease when the error reaches E_f, an apparent saturation value $E_f = 2.1 \cdot 10^{-2}E_\infty$.

A third stage of moderate growth follows, with a doubling time of approximately two days typical of synoptic scales and a slower decay of the growth rate. This phase that finally leads to saturation at E_∞ describes the growth of errors that initially were of the order of the analysis error as in the experiments by [Tribbia & Baumhefner, 2004].

In conclusion, even starting with band-limited or with very small initial errors, the structure and growth of forecast error affecting the quality of weather prediction is largely determined by baroclinic instability responsible for the development of synoptic weather systems.

3. Data Assimilation

In the Earth sciences the classical problem of estimating the state of a dynamical system from observations and the equations governing the systems evolution is known as data assimilation [Daley, 1991; Kalnay, 2003]. Data assimilation algorithms are used for diagnostic purposes and to provide the initial condition (the analysis) for weather prediction models. The analysis is obtained by optimally combining the information coming from a model forecast and the observations. Atmospheric and oceanic observations are noisy and very scattered in space and time. Dynamical models providing an approximate description of the evolution of the state of the atmosphere and the ocean are chaotic and, moreover, the number of degrees of freedom is huge ($> 10^9$).

The theory of data assimilation rests on the mathematical framework of estimation and optimal control theory [Jazwinski, 1970], but in Earth sciences we deal with complex chaotic models with a huge number of degrees of freedom that often defy straightforward application of classical methods.

The chaotic character of the real systems and of forecast models has two important implications. One is the need to provide an accurate estimate of the initial state to stretch the range of valuable forecasts towards the systems' predictability limits. The other, more subtle consequence of the divergence of trajectories is that the state estimation error itself strongly projects on the unstable manifold of the forecast model. Many authors have stressed the importance of the unstable manifold in data assimilation studies, but only recently the problem of estimating the state of a chaotic system has been systematically investigated.

Data assimilation has become a major area of research in meteorology and oceanography, as well as of numerical prediction. Based on estimation theory, advanced data assimilation methods have been

developed, the most notable examples in forecasting practice being four-dimensional variational assimilation (4DVar) [Sasaki, 1970; Le Dimet & Talagrand, 1986], and various schemes based on Kalman Filtering theory [Kalman, 1960] (see also [Ghil & Malanotte Rizzoli, 1991; Kalnay, 2003] and references therein). Depending on the particular application, many different schemes based on different approximations have been developed to cope with practical difficulties. We cannot cover here all the methods, all basically inspired by estimation theory, that have been applied, with varying level of success, in many realistic circumstances.

After a brief introduction of basic concepts and methods and in particular of 4DVar and of the Extended Kalman Filter, we present some recent work that addresses the problem of data assimilation in a chaotic system, based on a method referred to as Assimilation in the Unstable Subspace (AUS). The aim of this work is to investigate how the chaotic properties of a system affect the estimation problem and how to exploit this knowledge to improve the state estimate and, at the same time, to alleviate the computational burden needed to obtain it. The notion of unstable manifold, central topic of the present paper, plays a prominent role in the AUS approach.

3.1. *Basic concepts and definitions*

We consider the problem of estimating the state of a discrete dynamical system, that for the present purpose we assume is not affected by model error:

$$\mathbf{x}_{k+1} = \mathcal{M}_{k\to k+1}(\mathbf{x}_k), \quad k = 0, 1, 2, \ldots, \tag{16}$$

where the state at t_k is an n-vector. The distribution $\mathcal{P}(\mathbf{x}_0)$ of the initial condition is assumed given. Let discrete noisy observations $\mathbf{y}_k^o$ be given by:

$$\mathbf{y}_k^o = \mathcal{H}(\mathbf{x}_k) + \varepsilon_k^o, \quad k = 1, 2, \ldots, \tag{17}$$

where $\mathbf{y}_k^o$ is a p-vector and ε_k^o is a white Gaussian sequence, $\mathcal{N}(0, \mathbf{R}_k)$, independent of $\mathbf{x}_0$. Now let $\mathbf{Y}_l^o$ be the sequence of observations:

$$\mathbf{Y}_l^o = \mathbf{y}_1^o, \mathbf{y}_2^o, \ldots, \mathbf{y}_l^o. \tag{18}$$

Given a realization of the sequence of observations $\mathbf{Y}_l^o$, the problem consists in computing an estimate of $\mathbf{x}_k$. For $k < l$, $k = l$, $k > l$ the problem is referred to as smoothing, filtering and prediction, respectively.

In the probabilistic approach, the conditional probability density function (PDF) of $\mathbf{x}_k$:

$$\mathcal{P}(\mathbf{x}_k \,|\, \mathbf{Y}_l), \quad l = 1, 2, \ldots, k \tag{19}$$

given $\mathbf{Y}_l^o$ and $\mathcal{P}(\mathbf{x}_0)$ is the complete solution to the filtering problem. Given the correct initial PDF and assuming that the properties of the observational noise are known, the PDF's time evolution could in principle be predicted using the Fokker–Planck equation. In practice, in the fully nonlinear case it is impossible to determine the complete PDF. The fundamental difficulty of this approach lies in the high dimension of the state space, which makes it impossible to obtain the initial PDF, let alone compute its time evolution.

The problem simplifies enormously if the system's dynamics is linear. In the case of linear filtering and Gaussian errors the PDFs are characterized by the mean and the covariance matrices. By maximizing the conditional PDF one obtains the solution that minimizes the mean square error:

$$\mathcal{E}\{(\mathbf{x}_k - \mathbf{x}_k^t)^T \mathbf{S} (\mathbf{x}_k - \mathbf{x}_k^t)\} \tag{20}$$

where $\mathbf{x}^t$ is the *true* state, $\mathcal{E}\{\,\}$ is the expectation operator and $\mathbf{S}$ is a positive definite matrix. Under these conditions, the minimum error variance estimate is the conditional mean $\mathcal{E}\{\mathbf{x}_k^t \,|\, \mathbf{Y}_k\}$.

The Kalman Filter (KF), originally developed for linear systems propagates the state estimate together with the associated error covariance estimate given by

$$\mathbf{P}^a \equiv \mathcal{E}\{(\mathbf{x}_k - \mathbf{x}_k^t)(\mathbf{x}_k - \mathbf{x}_k^t)^T \,|\, \mathbf{Y}_k\}. \tag{21}$$

When observations become available, the KF solution provides the optimal linear estimate of the system's state [Kalman, 1960]; in this linear framework the propagation reduces to equations for the mean and covariance between observation times and the analysis update is given by:

$$\mathbf{x}_k^a = \mathbf{x}_k^f - \mathbf{K}_k \mathbf{H} \mathbf{x}_k^f + \mathbf{K}_k \mathbf{y}_k^o, \tag{22}$$

where $\mathbf{x}_k^f$ is the *forecast* obtained by integrating the model equations from a previous analysis, $\mathbf{H}$ is the linear observation operator and $\mathbf{K}_k$ is the Kalman gain matrix.

A different approach is the variational approach based on statistical (least square) estimation. The solution of the variational optimization problem is obtained by minimizing, with respect to $\mathbf{x}_0, \ldots, \mathbf{x}_K$, a cost function measuring the misfit

with observations and some *a-priori* estimate $\mathbf{x}_0^b$:

$$J = (\mathbf{x}_0 - \mathbf{x}_0^b)^T \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}_0^b) + \sum_{k=0}^{K} (\mathcal{H}(\mathbf{x}_k) - \mathbf{y}_k^o)^T \mathbf{R}_k^{-1} (\mathcal{H}(\mathbf{x}_k) - \mathbf{y}_k^o) \quad (23)$$

subject to the constraints given by the model equation (16). $\mathcal{H}$ is the nonlinear observation operator.

The positive definite matrices $\mathbf{B}^{-1}$, $\mathbf{R}^{-1}$ are to be regarded as *weighting* matrices, quantitative measures of our belief in the prior estimate, $\mathbf{x}_0^b$ and the observations. In 4DVar $\mathbf{x}_0^b$ can be a model forecast from a previous analysis as in the KF, the important difference being that in the KF its associated error covariance estimate $\mathbf{P}^f$ given by:

$$\mathbf{P}^f = \mathbf{M}\mathbf{P}^a\mathbf{M}^T, \quad (24)$$

that plays the role of $\mathbf{B}$, is also evolved from the previous analysis step. The statistical approach gives the same solution, the best linear unbiased estimate, as the probabilistic approach, provided the weighting matrices reflect the correct properties of the errors statistics.

The performance of 4DVar crucially depends on how good is the estimate of $\mathbf{B}$ that must be prescribed at $t = 0$. In fact, in practical applications of 4DVar, observations are collected in *batches* or *observation windows* to solve the problem separately for each *batch* of observations. Information from one batch is carried over to the next batch via the first term in the cost function. In atmospheric or oceanic applications, the model equations are nonlinear. Nonlinear estimation can be simplified by assuming that the error dynamics is linear: the Extended Kalman Filter (EKF) and 4DVar described in Secs. 3.4 and 3.5 are the most notable examples of sequential and variational algorithms of this sort.

3.2. *Data assimilation algorithms*

Data assimilation methods can be classified in two categories: sequential and variational, the most notable in the two classes being KF type schemes and 4DVar, respectively [Ghil & Malanotte Rizzoli, 1991; Kalnay, 2003] (and references therein).

The EKF is a straightforward way of extending the KF equations to the nonlinear case [Jazwinski, 1970; Miller *et al.*, 1994]: in the EKF, the error covariance is propagated in time according to linearized model dynamics. In the forecast step the nonlinear model is integrated, starting from the previous analysis to obtain the forecast state and the tangent linear equations propagate the analysis error covariance to obtain an estimate of the forecast error covariance. When observations become available they are assimilated through the analysis update that combines the forecast and the observations with the appropriate weights. The error covariance associated with the state estimate at the analysis step is also updated.

4DVar is an advanced technique that seeks the model trajectory that best fits the observations distributed within a given time interval with the dynamical constraint of the model equations [Talagrand & Courtier, 1987]. The adjoint technique [Le Dimet & Talagrand, 1986] allows the minimization of the 4DVar cost function to be made with respect to the state at the beginning of the interval. The solution of the minimization problem is obtained by forward integration of the model and backward integration of the adjoint of the tangent linear propagator that enters the expression of the cost function gradient. Given the same initial estimates of the background state and of the background error covariance, if errors are Gaussian and small enough to be described by the tangent linear equations, 4DVar and EKF give the same solution.

Some general comments are in place at this point. First of all it needs to be stressed that, in real world applications, the validity of the linear hypothesis depends on different factors: the characteristics of the observation network, the frequency and accuracy of observations, the rate of growth of errors, the properties of the assimilation scheme. In 4DVar, the length of the time window is another parameter that must be controlled. Second, what really matters is the asymptotic performance of the assimilation scheme. In Kalman filters, the error covariance matrix is updated at each step along the assimilation-forecast cycle and the asymptotic performance of the filter is independent of the guess of the error covariance at time $t = 0$. In addition, the evolved error covariance matrix in the Kalman Filter provides a measure of the goodness of the estimate. However, the computational cost of the error covariance propagation in the full EKF is prohibitively expensive for NWP models with order 10^9 degrees of freedom.

In 4DVar, the error covariance associated with the background state at the beginning of each assimilation window must be prescribed and the asymptotic performance depends, among other

factors, on the accuracy of this estimate. The reader is referred to [Kalnay, 2003] and [Tsuyuki & Miyoshi, 2007] for a review on the state of the art of data assimilation in meteorology; see also [Lorenc, 2003; Kalnay *et al.*, 2007; Gustafsson, 2007] for a discussion on the relative merits of 4DVar and Kalman Filters.

3.2.1. *Reducing the computational cost of the Kalman filter*

The problem of reducing the cost of the EKF in real applications has been addressed by a number of authors. In geophysical science literature reduced-rank approximations of the full EKF [Fukumori, 2002; Ghil & Malanotte Rizzoli, 1991; Tippet *et al.*, 2000; Todling & Cohn, 1994] confine the forecast error covariance matrix to a subspace of lower dimension by means of Singular Value Decomposition, Eigenvalue Decomposition or projection on the leading Empirical Orthogonal Functions [Pham *et al.*, 1998]. A Monte Carlo approach, referred to as Ensemble Kalman Filter (EnKF) [Evensen, 1994], provides ensemble representations for the probability distribution of the state estimate. The EnKF has also proven effective in reducing the computational cost associated with the full EKF. In the EnKF, the ensemble-predicted error statistics is estimated from the ensemble perturbations. In the forecast step, their evolution is computed as difference between nonlinear integrations of the model, a procedure similar to the breeding method [Toth & Kalnay, 1993, 1997]. If the analysis error is sufficiently small this procedure would give the same result as using the tangent linear propagator.

The various flavors of the EnKF testify the importance of this problem for real world applications (see [Blum *et al.*, 2008; Kalnay, 2003] and references therein). These include several refinements as perturbing observations, covariance localization, additive or multiplicative inflation. In particular, covariance localization is beneficial to prevent filter divergence when the number of ensemble members is small (for a review see [Evensen, 2003]). In all these studies, how many members of the ensemble are needed is estimated empirically on a case to case basis. Ott *et al.* [1995] estimate the local dimensionality (E-dimension) from the perturbations of the Local Ensemble Kalman Filter and the number of members of the ensemble necessary to represent the forecast error covariance in different geographical regions (tropics, extra-tropics, polar regions).

The Assimilation in the Unstable Space (AUS), developed by Trevisan and co-authors, consists in confining the analysis update in the subspace spanned by the leading unstable directions [Trevisan & Uboldi, 2004]. Applications to atmospheric and oceanic models [Uboldi & Trevisan, 2006; Carrassi *et al.*, 2008a] showed that even dealing with high-dimensional systems, an efficient error control can be obtained by monitoring only a limited number of unstable directions. More recently, Trevisan *et al.* [2010] formulated a reduced subspace four-dimensional assimilation algorithm, 4DVar-AUS (Four-dimensional Variational Assimilation in the Unstable Subspace). The key result of this study is the existence of an optimal subspace dimension for the assimilation that is approximately equal to the unstable and neutral subspace dimension. In recent work, [Trevisan & Palatella, 2010, 2011] investigated the consequences of the existence of an unstable manifold for the EKF solution. They presented a mathematically rigorous algorithm that solves the Kalman filter equations when all degrees of freedom are considered. They showed that the asymptotic rank of the EKF error covariance matrices is equal to the dimension of the unstable and neutral subspace. As a consequence, the full EKF and a reduced order algorithm (EKF-AUS) with assimilation increments limited to the unstable and neutral subspace were shown to produce the same results.

3.3. *The unstable manifold and data assimilation*

The nonlinear stability properties of the system do not only determine the predictability horizon of the initial value problem but also profoundly influence the assimilation process, affecting directly its quality and that of the subsequent forecast. If a system is chaotic, it is possible only by means of the observations, to keep the estimate from diverging from the true trajectory.

While all assimilation methods, more or less implicitly, exert some control on the flow dependent instabilities, algorithms referred to as Assimilation in the Unstable Subspace exploit the unstable subspace, the span of the leading Lyapunov vectors, as key dynamical information in the assimilation process.

The basic idea of confining the assimilation in the unstable subspace can be easily understood when errors behave linearly, that is when there are enough and sufficiently accurate observations that

the estimate remains at all times close to the true state. If the error at some time is small, its evolution can be described in terms of the evolution of the Lyapunov directions, with some projection coefficients. During the forecast, errors in the directions with positive (negative) Lyapunov exponents will amplify (decay). When the forecast is combined with the observations to form the analysis, error components that have amplified will be reduced; only if the observations are perfect, these components can be reduced to zero. Forecast errors in decaying directions naturally tend to zero if observations are not used or are used with the proper weight. By confining the analysis increment only in the unstable and neutral subspace, with the AUS approach, it is thus possible to keep the estimated trajectory close to the true trajectory, at least in a perfect model setting.

The first applications of the AUS assimilation dealt with the problem of selecting the locations, that depend on the particular weather situation, of supplementary observations meant to achieve the greatest improvement in analysis and forecasts. This problem was one of the research challenges undertaken by the international program of THORPEX (The Observing system Research and Predictability Experiment) to accelerate the improvement in high impact weather forecast: additional observations could be *adaptively* located in geographical regions where the fixed observations were not sufficient to control the growth of errors due to flow dependent instabilities.

The error in an assimilation-forecast cycle is expected to depend on the structure, growth, and propagation properties of flow instabilities, as well as on the temporal and geographical distribution of observations. These errors can be most efficiently reduced by tracking the structure of the errors in the most unstable directions that develop along the trajectory and by placing observations in the locations where the maximum amplitude of the structure is attained. By confining the analysis increment in the subspace $\mathbf{E}$ spanned by the most unstable vectors, the analysis solution reads:

$$\mathbf{x^a} = \mathbf{x^b} + \mathbf{E\Gamma E}^T\mathbf{H}^T(\mathbf{R} + \mathbf{HE\Gamma E}^T\mathbf{H}^T)^{-1} \times (\mathbf{y}^o - \mathbf{Hx^b}) \tag{25}$$

where the background covariance matrix $\mathbf{B} = \mathbf{E\Gamma E}^T$ [Trevisan & Uboldi, 2004]. The most efficient and *safe* (for reasons related to the stability of the solution) way of exploiting just a few unstable directions is to use AUS in combination with observations located where each column of $\mathbf{E}$ attains its maximum value.

3.4. *Four-dimensional variational assimilation and its reduction to the unstable subspace (4DVar-AUS)*

4DVar generates a model trajectory that best fits the observations available within a given assimilation window. Within the assimilation window, the flow dependent instabilities are naturally described in 4DVar by the tangent linear propagator. In addition, at the start of each assimilation window, an *a-priori* estimate of the background error covariance is needed [Bannister, 2008].

For long assimilation windows, 4DVar analysis errors are known to project on the unstable subspace of the system [Pires *et al.*, 1996]. Errors in the stable directions that would be damped in the long range, for short assimilation windows are non-negligible in the analysis and affect the next assimilation cycle, causing short term enhanced error growth as noticed by [Swanson *et al.*, 2000] and discussed by [Trevisan *et al.*, 2010]. It is possible to avoid introducing errors in the stable directions if we confine the analysis increment in the unstable and neutral subspace of the system; this is obtained by applying to 4DVar the AUS assimilation constraint. The dynamical information on the growth of errors in the unstable and neutral directions, the Lyapunov vectors with positive and null exponents, is explicitly estimated in the 4DVar-AUS algorithm [Trevisan *et al.*, 2010] and the adjoint integration is not needed.

3.4.1. *The 4DVar algorithm*

Strong constraint 4DVar seeks the (nonlinear) best estimate of the initial state $\mathbf{x}_0$ which minimizes the misfit with observations available in a given time interval (assimilation window) and with a background state $\mathbf{x}_0^b$. The standard cost function for strong constraint 4DVar, in discrete form, can be written as:

$$J(\mathbf{x}_0) = (\mathbf{x}_0 - \mathbf{x}_0^b)^T\mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) + \sum_{k=0}^{K}(\mathcal{H}_k(\mathbf{x}_k) - \mathbf{y}_k^o)^T\mathbf{R}^{-1}(\mathcal{H}_k(\mathbf{x}_k) - \mathbf{y}_k^o) \tag{26}$$

where $\mathbf{y}_k^o$ are the observations available at discrete times $t_k = k\Delta t, k = 0, \ldots, K$, within the assimilation window of length $\tau = t_K - t_0$; $\mathbf{B}$ and $\mathbf{R}$ represent the background and observation error covariance matrices, $\mathcal{H}$ the nonlinear observation operator, and the sequence of model states $\mathbf{x}_k$ is a solution of the nonlinear model equations:

$$\mathbf{x}_k = \mathcal{M}_{0\to k}(\mathbf{x}_0). \tag{27}$$

The control variable for the minimization is the model state $\mathbf{x}_0$ at the beginning of the assimilation window. Given the tangent linear equations describing the evolution of infinitesimal perturbations $\delta\mathbf{x}_k$ relative to an orbit of Eq. (16):

$$\delta\mathbf{x}_k = \mathbf{M}_{0\to k}\delta\mathbf{x}_0, \tag{28}$$

the gradient of J with respect to $\mathbf{x}_0$ can be written as:

$$\frac{1}{2}\nabla_{\mathbf{x}_0} J = \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) + \sum_{k=0}^{K} \mathbf{M}_{0\to k}^T \mathbf{H}_k^T \mathbf{R}^{-1}(\mathcal{H}_k(\mathbf{x}_k) - \mathbf{y}_k) \tag{29}$$

where $\mathbf{H}_k$ represents the linearized observation operator, and the superscript T stands for transpose.

For a given nonlinear trajectory, the gradient can be estimated by use of the adjoint method. The solution of the minimization problem is obtained by forward integration of the model and backward integration of the adjoint, with an iterative descent algorithm.

3.4.2. *The 4DVar-AUS algorithm*

The 4DVar-AUS algorithm consists in confining the increment $\delta\mathbf{x}_0$ which minimizes the cost function in the reduced dimension subspace spanned by the N most unstable directions of the system corresponding to the leading N Lyapunov exponents. When $N = n$ the full space standard 4DVar solution is recovered.

Let $\mathbf{E}_0$ be the matrix whose columns are the orthonormal tangent vectors spanning the N-dimensional most unstable subspace of the system at t_0. The linear evolution within the assimilation window $[t_0, \tau]$ is given by:

$$\mathbf{M}_{0\to k}\mathbf{E}_0 = \mathbf{E}_k\mathbf{\Lambda}_k, \quad k = 0, \ldots, K \tag{30}$$

where

$$\mathbf{\Lambda}_k = \operatorname{diag}\left[\Lambda_k^{(1)}, \Lambda_k^{(2)}, \ldots, \Lambda_k^{(N)}\right] \tag{31}$$

and $\Lambda_k^{(j)}$ is the jth local amplification factor. The Gram–Schmidt orthonormalization is applied at $t = t_0$. Let the increment $\delta\mathbf{x}_0$ be confined in the subspace $\mathbf{E}_0$ and its projection $\widetilde{\delta\mathbf{x}_0}$ be given by:

$$\widetilde{\delta\mathbf{x}_0} = \mathbf{E}_0\mathbf{E}_0^T\delta\mathbf{x}_0 \tag{32}$$

where for simplicity, the Euclidean norm is adopted. The evolution of the projected increment is governed by:

$$\begin{aligned}\widetilde{\delta\mathbf{x}_k} &= \mathbf{M}_{0\to k}\mathbf{E}_0\mathbf{E}_0^T\delta\mathbf{x}_0 \\ &= \mathbf{E}_k\mathbf{\Lambda}_k\mathbf{E}_0^T\delta\mathbf{x}_0, \quad k = 0, \ldots, K.\end{aligned} \tag{33}$$

The cost function gradient in the reduced subspace can be written as:

$$\frac{1}{2}\widetilde{\nabla_{\mathbf{x}_0} J} = \mathbf{E}_0\left[\mathbf{E}_0^T\mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) + \sum_{k=0}^{K}\mathbf{\Lambda}_k\mathbf{E}_k^T\mathbf{H}_k^T\mathbf{R}^{-1}(\mathcal{H}_k(\mathbf{x}_k) - \mathbf{y}_k)\right]. \tag{34}$$

The minimization proceeds as in the 4DVar algorithm (for details see [Trevisan *et al.*, 2010]) and in the assimilation cycle the estimated state and its associated unstable subspace are used to initialize the next assimilation window. Notice that no use of the adjoint integration is made. In the 4DVar-AUS assimilation, N is (at least) equal to the number of positive and null Lyapunov exponents.

3.4.3. *4DVar-AUS results*

To illustrate the behavior of the algorithm we use the [Lorenz, 1996] model. Numerical experiments were performed with three different configurations (40, 60 and 80 longitudinal grid points). The number of positive exponents (13, 19, 26) in the three different systems is proportional to the spatial extension of the domain. The unstable and neutral subspace, the span of Lyapunov vectors with positive or null exponents contains the most rapidly growing perturbations. 4DVar-AUS confines the assimilation of observations in this subspace. In standard 4DVar, errors at the beginning (end) of the assimilation window are confined in the stable (unstable) subspace [Pires *et al.*, 1996]. In 4DVar-AUS, after a certain number of iterations of the

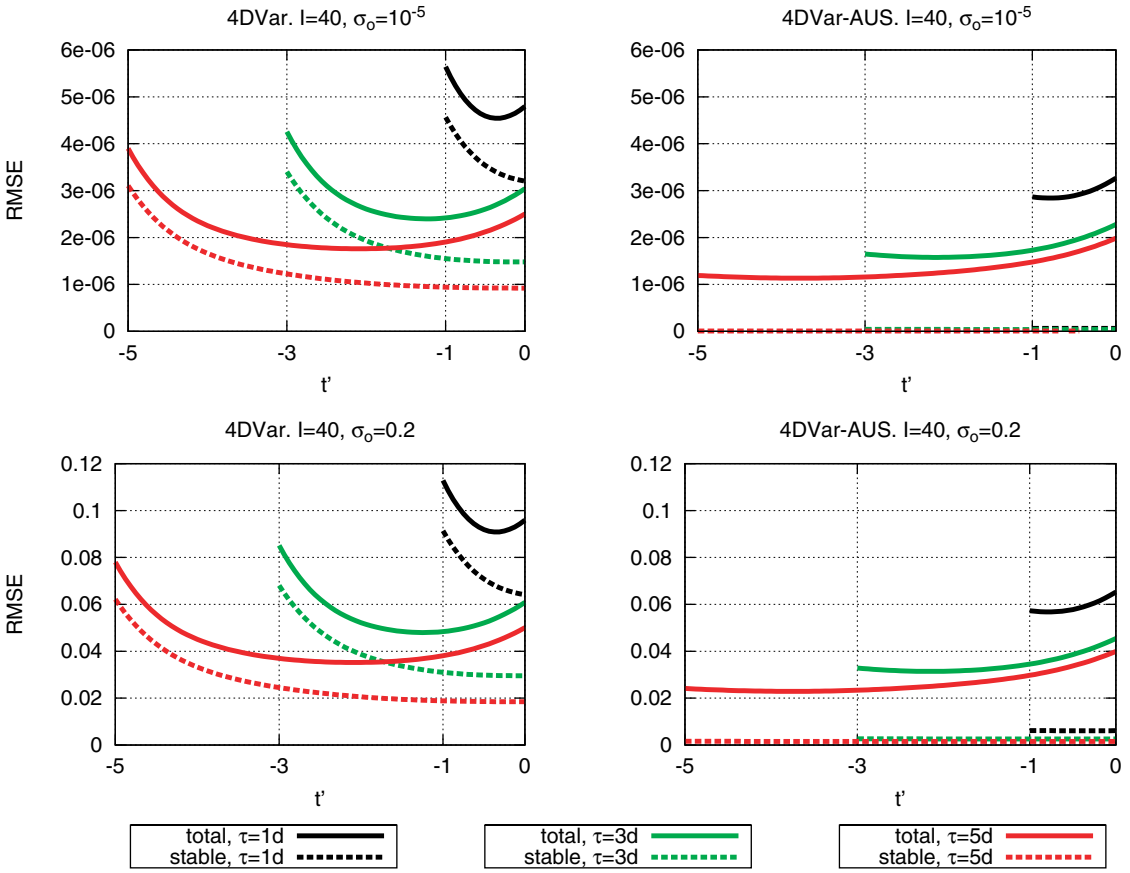

Fig. 8. After [Trevisan *et al.*, 2010]. Time average RMS error within $1, 3, 5$ days assimilation windows as a function of $t' = t - \tau$, with $\sigma_o = 0.2, 10^{-5}$ for the model configuration $n = 40$. Left panel: 4DVar. Right panel: 4DVar-AUS with $N = 15$. Solid lines refer to total assimilation error, dashed lines refer to the error component in the stable subspace.

assimilation cycle, errors are confined in the unstable subspace both at the beginning and at the end of the assimilation window (see Fig. 8).

Furthermore, the error of 4DVar-AUS is smaller than the error of 4DVar, particularly for short assimilation windows, when the errors in the stable directions do not have enough time to be damped and affect the 4DVar solution also at the end of the window. One of the main results of the study was to find that, in the presence of observational error, there is an optimal subspace dimension for the assimilation that is approximately equal to $N^+ + N^0$, where N^+ and N^0 are the number of positive and null Lyapunov exponents (see Fig. 9).

The 4DVar solution that minimizes the cost function in the whole space ($N = n$), while closer to the observations, is farther away from the truth. This result has been explained showing that, when assimilating in the unstable and neutral subspace, errors in the stable directions are naturally damped. Because of observational error, assimilating in the whole space otherwise keeps the stable components of the error alive, deteriorating the overall assimilation performance. If, instead, the observational error is zero, the optimal dimension is the dimension n of the whole space.

3.5. *The Extended Kalman Filter and its reduction to the Unstable Subspace (EKF-AUS)*

In the EKF, the propagation of the flow dependent instabilities is obtained by explicitly evolving the analysis error covariance estimate from the previous analysis step: the dimension of the error covariance matrices is equal to the state dimension, n. In a reduced form, referred to as EKF-AUS, the assimilation is performed in the unstable and neutral subspace of the tangent space, of dimension m and the rank of the error covariance matrices is equal to m. This subspace, the span of the Lyapunov vectors with positive and null exponents is the subspace where instabilities develop along the system's trajectory and usually has a dimension much smaller

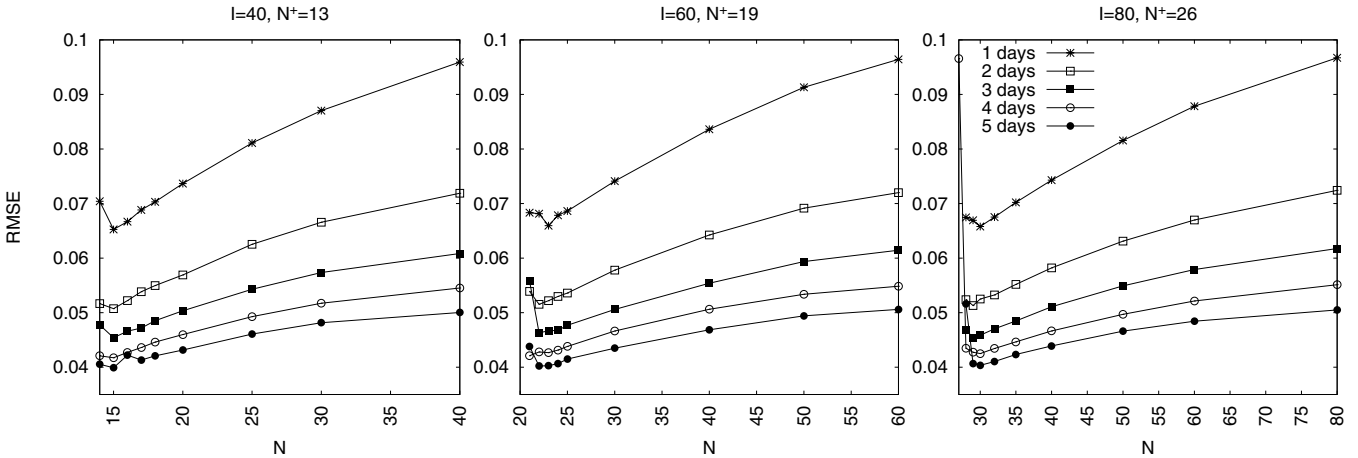

Fig. 9. After [Trevisan *et al.*, 2010]. Time average RMS analysis error at $t = \tau$ as a function of the subspace dimension N for three model configurations: $n = 40, 60, 80$. Different curves in the same panel refer to different assimilation windows from 1 to 5 days. The observation error standard deviation is $\sigma_o = 0.2$.

than the total number of degrees of freedom. For instance, Carrassi *et al.* [2008a, 2008b] found that in a quasi-geostrophic model with 14784 degrees of freedom, the number m of positive exponents is equal to 25.

As shown in [Trevisan & Palatella, 2010, 2011], EKF-AUS not only reduces the cost of the full EKF, but it leads to the same solution.

3.5.1. *The EKF algorithm*

In the EKF, the state evolves according to the full nonlinear equations and the tangent linear operator is used to predict the approximate error statistics. The estimate of the state, referred to as the analysis $\mathbf{x}^a$, is obtained by combining a forecast state $\mathbf{x}^f$ with the possibly incomplete and noisy observations $\mathbf{y}_k^o = \mathcal{H}(\mathbf{x}_k) + \varepsilon_k^o$ given at discrete times $t_k \geq t_0$, $k \in \{1, 2, \ldots\}$. The observation error ε^o is assumed to be Gaussian with zero mean and known covariance matrix $\mathbf{R}$. $\mathcal{H}$ is the observation operator. The estimate update is given by the analysis equation:

$$\mathbf{x}_k^a = \mathbf{x}_k^f - \mathbf{K}_k \mathcal{H}(\mathbf{x}_k^f) + \mathbf{K}_k \mathbf{y}_k^o, \tag{35}$$

where $\mathbf{x}_k^f$ is the *forecast* obtained by integrating the nonlinear equations from a previous analysis time:

$$\mathbf{x}_k^f = \mathcal{M}(\mathbf{x}_{k-1}^a), \tag{36}$$

$\mathcal{M}$ being the nonlinear evolution operator. $\mathbf{K}_k$ is the *gain matrix* at time t_k given by:

$$\mathbf{K}_k = \mathbf{P}_k^f \mathbf{H}^T \left(\mathbf{H}\mathbf{P}_k^f \mathbf{H}^T + \mathbf{R}\right)^{-1}, \tag{37}$$

where $\mathbf{H}$ is the Jacobian of $\mathcal{H}$. The analysis error covariance update equation is given by:

$$\mathbf{P}_k^a = (\mathbf{I} - \mathbf{K}_k \mathbf{H}) \mathbf{P}_k^f, \tag{38}$$

and $\mathbf{P}_k^f$, the forecast error covariance, is given by:

$$\mathbf{P}_k^f = \mathbf{M}_k \mathbf{P}_{k-1}^a \mathbf{M}_k^T, \tag{39}$$

where $\mathbf{M}$ is the linearized evolution operator associated with $\mathcal{M}$.

3.5.2. *The EKF-AUS algorithm*

The algorithm belongs to the family of square-root implementations of the Extended Kalman Filter [Thornton & Bierman, 1976; Hamill & Whitaker, 2011]. The assimilation is performed in a manifold of dimension m. When m is equal to the total number of degrees of freedom of the system, the algorithm solves the standard EKF equations. When $m = N^+ + N^0$, where N^+ and N^0 are the number of positive and null Lyapunov exponents, the reduced form, with Assimilation in the Unstable Subspace (EKF-AUS) is obtained.

At time $t = t_{k-1}$ (here and in the following we drop the time-step subscript from the equations since, unless otherwise stated, all terms refer to the same time step k), let the columns of the $n \times m$ matrix $\mathbf{X}^a = [\mathbf{x}_1^a, \mathbf{x}_2^a, \ldots, \mathbf{x}_m^a]$ be orthogonal, with $\mathbf{X}^a$ one of the square roots of $\mathbf{P}^a$, namely:

$$\mathbf{P}^a = \mathbf{X}^a \mathbf{X}^{aT}. \tag{40}$$

At time $t = t_0$, the vectors $\mathbf{x}_i^a$, $i = 1, 2, \ldots, m$ are arbitrary independent initial perturbations.

In the standard EKF algorithm, the number of perturbations is equal to the total number of degrees of freedom of the system, $m = n$.

3.5.2.1. The forecast step

In the forecast step, the tangent linear operator $\mathbf{M}$ acts on the perturbations $\mathbf{X}^a$ defined at (analysis) time $t = t_{k-1}$: $\mathbf{X}^f = \mathbf{M}\mathbf{X}^a$ where $\mathbf{X}^f = [\mathbf{x}_1^f, \mathbf{x}_2^f, \ldots, \mathbf{x}_m^f]$. The $n \times n$ forecast error covariance matrix, $\mathbf{P}^f$ can be cast in the form:

$$\mathbf{P}^f = \mathbf{X}^f \mathbf{X}^{fT} \equiv \mathbf{E}^f \mathbf{\Gamma}^f \mathbf{E}^{fT} \tag{41}$$

where the m columns of $\mathbf{E}^f$ are obtained by a Gram–Schmidt orthonormalization of the columns of $\mathbf{X}^f$. The $m \times m$ (in general nondiagonal) symmetric matrix $\mathbf{\Gamma}^f$ defined as:

$$\mathbf{\Gamma}^f = \mathbf{E}^{fT} \mathbf{X}^f \mathbf{X}^{fT} \mathbf{E}^f \tag{42}$$

represents the forecast error covariance matrix, confined to the subspace $\mathcal{S}^m$ of the evolved perturbations.

3.5.2.2. The analysis step

Using the definition of $\mathbf{P}^f$ of Eq. (41) the Kalman gain expression becomes:

$$\mathbf{K} = \mathbf{E}^f \mathbf{\Gamma}^f (\mathbf{H}\mathbf{E}^f)^T [(\mathbf{H}\mathbf{E}^f) \mathbf{\Gamma}^f (\mathbf{H}\mathbf{E}^f)^T + \mathbf{R}]^{-1} \tag{43}$$

and the usual analysis error covariance update Eq. (38) reads:

$$\begin{aligned} \mathbf{P}^a &= (\mathbf{I} - \mathbf{K}\mathbf{H}) \mathbf{E}^f \mathbf{\Gamma}^f \mathbf{E}^{fT} \\ &= \mathbf{E}^f \mathbf{\Gamma}^f \mathbf{E}^{fT} - \mathbf{E}^f \mathbf{\Gamma}^f (\mathbf{H}\mathbf{E}^f)^T \\ &\quad \times [(\mathbf{H}\mathbf{E}^f) \mathbf{\Gamma}^f (\mathbf{H}\mathbf{E}^f)^T + \mathbf{R}]^{-1} \mathbf{H}\mathbf{E}^f \mathbf{\Gamma}^f \mathbf{E}^{fT} \\ &= \mathbf{E}^f \{\mathbf{\Gamma}^f - \mathbf{\Gamma}^f \mathbf{E}^{fT} \mathbf{H}^T \\ &\quad \times [(\mathbf{H}\mathbf{E}^f) \mathbf{\Gamma}^f (\mathbf{H}\mathbf{E}^f)^T + \mathbf{R}]^{-1} \mathbf{H}\mathbf{E}^f \mathbf{\Gamma}^f \} \mathbf{E}^{fT} \\ &\equiv \mathbf{E}^f \mathbf{\Gamma}^{a'} \mathbf{E}^{fT}. \end{aligned} \tag{44}$$

The analysis error covariance matrix $\mathbf{P}^a$ can be written as:

$$\begin{aligned} \mathbf{P}^a &= \mathbf{E}^f \mathbf{\Gamma}^{a'} \mathbf{E}^{fT} \\ &= \mathbf{E}^f \mathbf{U} \mathbf{\Gamma}^a \mathbf{U}^T \mathbf{E}^{fT} \\ &\equiv \mathbf{E}^a \mathbf{\Gamma}^a \mathbf{E}^{aT}, \end{aligned} \tag{45}$$

where the columns of the $(m \times m)$ orthogonal invertible matrix $\mathbf{U}$ are the eigenvectors of the symmetric matrix $\mathbf{\Gamma}^{a'} = \mathbf{U}\mathbf{\Gamma}^a\mathbf{U}^T$ and $\mathbf{\Gamma}_a = \mathrm{diag}[\gamma_i^2]$ is diagonal. Therefore, the m columns of $\mathbf{E}^a = [\mathbf{e}_1^a, \mathbf{e}_2^a, \ldots, \mathbf{e}_m^a]$ obtained by

$$\mathbf{E}^a = \mathbf{E}^f \mathbf{U} \tag{46}$$

span the same subspace $\mathcal{S}^m$ as the columns of $\mathbf{E}^f$. Consequently, the analysis step preserves the subspace $\mathcal{S}^m$, an important point on the comparison of our scheme with the Benettin *et al.* algorithm [Benettin *et al.*, 1980]. The square root of $\mathbf{P}^a$, written as $\mathbf{P}^a = \mathbf{E}_a \mathbf{\Gamma}_a \mathbf{E}_a^T = \mathbf{X}_a \mathbf{X}_a^T$, provides a set of orthogonal vectors:

$$\mathbf{X}^a = \mathbf{E}^a (\mathbf{\Gamma}^a)^{1/2}. \tag{47}$$

The columns of

$$\begin{aligned} \mathbf{X}^a &= [\gamma_1 \mathbf{e}_1^a, \gamma_2 \mathbf{e}_2^a, \ldots, \gamma_m \mathbf{e}_m^a] \\ &= [\mathbf{x}_1^a, \mathbf{x}_2^a, \ldots, \mathbf{x}_m^a] \end{aligned} \tag{48}$$

are the new set of perturbation vectors defined after the analysis step at time t_k that enter the forecast step at the next time step t_{k+1}. When $m = n$, $\mathbf{K}$ is the usual Kalman gain. In EKF-AUS ($m = N^+ + N^0$) the analysis increment is confined to the subspace $\mathcal{S}^m$ spanned by the m columns of $\mathbf{E}^f$ in view of the form of $\mathbf{K}$.

Trevisan and Palatella [2011, 2010] compared the steps in the evolution of perturbations in the EKF algorithm with those of [Benettin *et al.*, 1980] procedure to compute Lyapunov vectors. In the assimilation cycle, the analysis step preserves the subspace and reduces the amplitude of the perturbations while the forecast perturbations $\mathbf{X}^f$ are repeatedly orthogonalized but not orthonormalized. Thus, in the long run perturbations in the stable directions will decay and only the unstable and neutral subspace Lyapunov vectors will, in the end, be the same as those computed in the [Benettin *et al.*, 1980] algorithm. Therefore the EKF and EKF-AUS algorithms will asymptotically produce the same estimate of error covariances and the two algorithms will be equivalent.

3.5.3. *EKF-AUS results*

As in the 4DVar-AUS results section, to illustrate the behavior of the algorithm we use the [Lorenz, 1996] model. To illustrate the properties of the EKF and EKF-AUS solutions, numerical experiments were performed with three different configurations (40, 60 and 80 longitudinal grid points). The numbers of positive exponents are 13, 19, 26, respectively.

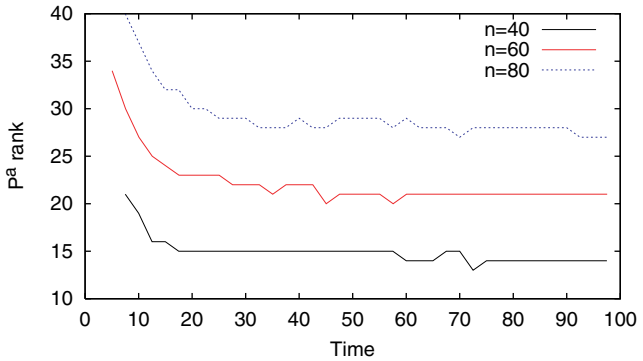

Fig. 10. After [Trevisan & Palatella, 2011]. Rank of the analysis error covariance matrix $\mathbf{P}^a$ in the EKF algorithm. For $n = 40, 60, 80$, the dimensions of the unstable and neutral space are $14, 20, 26$, respectively. $\sigma_0 = 0.01$.

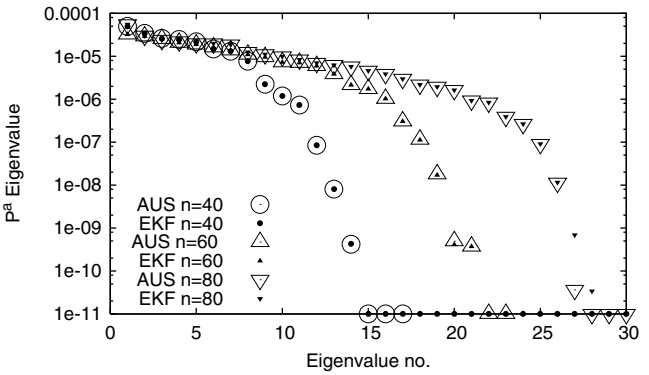

Fig. 12. After [Trevisan & Palatella, 2011]. Eigenvalues of the analysis error covariance matrix at final time of assimilation with: EKF (full) and EKF-AUS (empty) for systems with $n = 40, 60, 80$. $\sigma_0 = 0.01$.

The numerical experiments confirm the theoretical prediction that the rank of $\mathbf{P}^a$ in the full EKF decays in time and asymptotically becomes as small as the dimension of the number of positive and null Lyapunov exponents of the original system (Fig. 10). This is because only the estimated errors in the unstable and neutral subspace survive the filtering process.

Another key result confirmed by the numerical experiments is that the full EKF and its reduced form EKF-AUS give the same results: the average rms error (Fig. 11) and the eigenvalues of the asymptotic analysis covariance matrix (Fig. 12) are in fact the same.

It should be stressed that all considerations regarding the behavior of the error estimate are valid in the framework of EKF theory based on the assumption that errors behave linearly. If errors do not behave linearly, the EKF will still provide the same error estimate, but the true error may not be confined in the unstable and neutral subspace but live in a higher-dimensional subspace. In that case, in order to avoid filter divergence, it may be necessary to artificially augment the subspace of the estimated error.

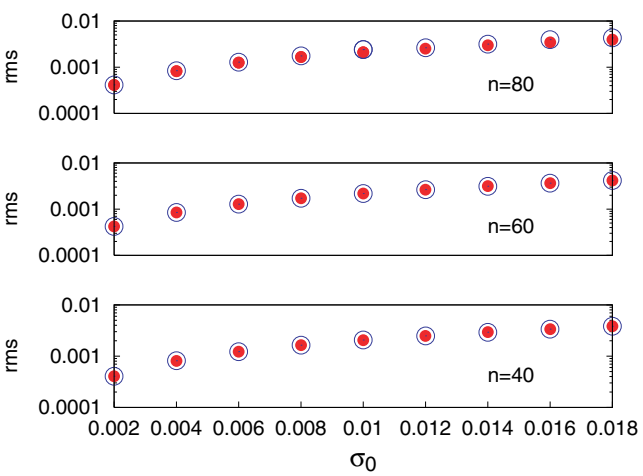

Fig. 11. After [Trevisan & Palatella, 2011]. Average root mean square error of the algorithms: EKF (full circles), EKF-AUS (empty circles) for systems with $n = 40, 60, 80$ degrees of freedom. The average is performed over 1000 assimilation steps.

4. Model Error as an Additional Source of Uncertainty

So far, the predictability problem has been discussed in the *perfect model* framework, by assuming that the equations governing the system's evolution are perfectly known. However, forecast models are but a discrete and crude approximation of the underlying *true* system dynamics; inevitable uncertainties are associated with the numerical formulation of the dynamical equations, including the influence of unresolved scales. *Model errors* are the result of parameterized, unresolved, unmodeled or stochastic processes.

The effect of model error in weather forecasting is a very difficult and complex topic and several authors in the last years are devoting their efforts to this argument. For this reason we will briefly describe the main points of model error theory and its connection to the theory of chaotic systems.

If the *real* system that we want to model is chaotic, the effect of model imperfections is particularly critical, whether we use the model for forecasting or diagnostic purposes. The model's attractor can only be an approximation of the true system's attractor so that the climate (the

statistical properties of the atmospheric variables) of the model and the real climate cannot be exactly the same. In forecasting the weather, as any other chaotic system, model errors add up to initial condition errors at each time step along the integration of the evolution equations and take part in the same amplification process. We have stressed the importance of Lyapunov vectors for the description of preferred directions of error amplification in predictability and assimilation problems. In the presence of model error, even the dominant Lyapunov vectors and exponents of the model cannot be exactly the same as those of the real system, one reason being that the model cannot capture the unresolved scale instabilities.

Whereas systematic error is due to model imperfections, nonsystematic errors are due to the combination of initial condition and model error. Roughly speaking, assuming that the analysis error is known, model error could in principle be estimated by subtracting the predictability error from the forecast error. However, even assuming that error growth is well reproduced by the model at least in the resolved scales, in order to disentangle the different — initial condition and model — error sources of forecast error, the analysis error and the forecast error need to be accurately estimated. A useful step would be the construction of ensemble forecasting systems with the property of reproducing the correct distribution of predictability error [Candille & Talagrand, 2005].

From the practical perspective, the adverse effects of model error have been reduced through the never ending process of improving models by increasing resolution, introducing new physics and more efficient parameterizations and resolving other models' shortcomings by model calibration, i.e. *a-posteriori* corrections of systematic error. Verification studies provide evidence that improvements in forecast skill have been obtained due to the reduction of both initial conditions' and of models' systematic errors. A number of techniques have been developed in order to deal with model uncertainty in ensemble forecasting: stochastically perturbed parameterization tendencies [Buizza *et al.*, 1999], stochastic backscatter [Berner *et al.*, 2009] and multimodel or multicenter ensembles approaches ([Bougeault *et al.*, 2010] and references therein).

Model error is often represented as a stochastic noise ϵ_k. The explicit assumption of no time correlation is made in standard Kalman Filtering theory by adding a model error covariance term $Q_k = \langle \epsilon_k^q \epsilon_k^{qT} \rangle$ to the forecast error covariance $\mathbf{P}^f$. Daley [1992] examined the effect of time correlated model error in the framework of Kalman filtering. Recent studies have shown that treating model error as part of the estimation problem leads to significant improvements in the accuracy of the state estimate [Vidard *et al.*, 2004; Li *et al.*, 2009]. Tremolet [2007] presented a formulation of weak-constraint 4DVar that explicitly represents model error as part of the 4DVar control variable.

A method of estimating and accounting for model error in the context of an ensemble Kalman filter technique was developed in [Mitchell & Houtekamer, 2000]. The method involves parameterizing the model error and using innovations to estimate the model-error parameters. Model bias or climate drift is another effect of model error. The general problem of model bias estimation in Kalman Filtering was first studied in [Friedland, 1969]. A statistical method was originally proposed in [Leith, 1978] to account for model bias and systematic errors linearly dependent on the flow anomalies. Inspired by the papers by Leith [1978], DelSole and Hou [1999], Danforth *et al.* [2007] proposed a procedure suitable for operational use at relatively small computational expense to estimate and reduce state-dependent model systematic errors. The method requires a time series of analysis increments (corrections) to estimate the error covariance.

With the above techniques, model bias is corrected by moving the state estimate from the model attractor to the system attractor. A different approach is that of searching for a state estimate that best represents the true state on the model attractor. To reduce drift-induced errors Toth and Peña [2007] introduced a mapping vector to move observations from near the attractor of nature to the vicinity of the model attractor before performing data assimilation.

The deterministic and probabilistic aspects of the model error dynamics has been analyzed in a series of studies [Nicolis, 2003, 2004]. Some generic features have been identified: the signature of the deterministic nature of the model error evolution is found to markedly affect the short-time regime of the error evolution and, furthermore, the role of the Lyapunov exponents on the model error dynamics is revealed. By using a time series expansion of the error dynamics, it is found that the short-time regime of the error variance is bound to be

quadratic in time, the spectrum of Lyapunov exponents is related to the duration of this short-time regime [Nicolis, 2003] and, in the case of errors arising from the presence of unresolved scales, to the first higher order [Nicolis, 2004]. This approach has been exploited in the context of sequential assimilation [Carrassi *et al.*, 2008b] and to formulate a weak-constraint 4DVar by deriving a short-time approximation for the model error correlations.

More recently Nicolis *et al.* [2009] studied the case of small uncorrelated error in the initial condition and in the model parameters. They found the existence of a minimum for the error variance as well as the time interval at which the contribution of the model and initial condition error are equal. The ability and efficiency of using Markov Gaussian noise to represent and correct model error in the long term prediction has been studied by the same author [Nicolis, 2005]. It has been found that this approach is not always adequate as different responses to the noise can be obtained depending on the structure of the model equations.

In any case, it is important to point out that the topic of the influence of model error in chaotic systems and its interaction with the initial condition errors is very complex. Scientists working in this field still have to face several problems both from theoretical and practical points of view. We address the reader interested on the topic of model error to an article explicitly devoted to this theme in the same issue of this journal.

5. Discussion and Further Developments

The atmospheric community has long been aware that understanding error growth dynamics and estimating the predictability properties of atmospheric flow at various scales is important for effectively designing, observing and forecasting systems and for estimating the forecast skill. A notable example is how predictability theory helped construct current Ensemble Prediction Systems able to predict flow-dependent uncertainties in medium range forecasts.

The growth of errors in atmospheric models is governed by the instabilities that characterize chaos. In the present paper, the accent has been placed throughout on the importance of the Lyapunov vectors to describe the growth of prediction errors at different time and space scales. A fundamental property of the unstable manifold is its variation with the flow: the unstable structures that develop in time can be estimated by computing the Lyapunov vectors. This can be done by integrating the tangent linear equations along the trajectory of the model that simulates the evolution of the atmospheric flow or by the breeding method. A set of N initial random perturbations, subject to repeated orthonormalization, after a transient time, will be oriented in the N directions spanning the most unstable subspace; in reviewing established theory we have seen how this subspace will, thereafter be preserved along the flow.

Another important property is the relation with the local geometry of the attractor; if the simulated trajectory is sufficiently close to the true trajectory, the difference of the corresponding states lies on the unstable (and neutral) manifold. Therefore, given a sufficiently accurate estimate of the true state and of the unstable tangent subspace one knows that the true state lies somewhere in this subspace, of usually much lower dimension than the total phase space.

This property turned out to be very important in the problem of estimating the state of a system, given its evolution equations and noisy observations of the true state. When correcting a model state by assimilating observations, the estimate of the unstable directions can be used to constrain the analysis increment in their subspace. This constraint is at the basis of data assimilation algorithms that are referred to as AUS. In the hypothesis that the pseudo-trajectory, the solution of the estimation problem, remains close enough to the true trajectory, Lyapunov exponents and vectors, that are well defined for the true trajectory, are well approximated by the corresponding quantities computed along the estimated trajectory. In this case, the assimilation algorithms with confinement in the unstable subspace give results that are as good (EKF-AUS), in the sense of minimizing the error, or better (4DVar-AUS) than assimilating in the whole space. This result is based on the assumption that errors behave linearly and it may not apply in many realistic circumstances, but it is valid whenever observations are sufficiently dense and accurate that the estimate remains at all times close to the true trajectory.

A by-product of the assimilation in the unstable subspace is the estimate of the directions spanning the unstable manifold. An estimate of the unstable subspace dimension can be obtained from the asymptotic assimilation solution, given an

overestimate for its value (that can be much smaller than the total number of degrees of freedom) used as an initial guess. In the EKF, the estimate of the dimension of the unstable subspace is given by the rank of the asymptotic error covariance matrix. In 4DVar-AUS it is the equal to the optimal (in the sense of minimizing the error) dimension for the confinement.

It is worth discussing at this point why the 4DVar-AUS algorithm turns out to be superior to standard 4DVar, while EKF-AUS gives the same results as the standard EKF. The interpretation is simple. The full EKF performs exactly as EKF-AUS because the estimated asymptotic error covariances are the same. In this sense, 4DVar-AUS, that confines the analysis increment in the unstable and neutral subspace is equivalent to the EKF. Standard 4DVar instead looks for the minimizing solution in the full space; when the error dynamics is linear, 4DVar-AUS is superior to 4DVar because it looks for the solution in the manifold where the true state is found.

When the error dynamics is not linear, the situation is different. In 4DVar-AUS the dimension of the subspace for the confinement is kept fixed along the assimilation cycle and is the same at the beginning of each time window. (Within a given window, of limited time extension, the independent orthogonal directions are unlikely to collapse to a subspace of lower dimension.)

In EKF and EKF-AUS assimilation, instead, the cyclic evolution of the estimated error is linear so that, asymptotically the error estimate and the analysis increment is always confined in the unstable (and neutral) subspace. However, if nonlinearity is important, the true error will not remain exactly in the unstable and neutral subspace. All Kalman type filters, where the estimated error subspace is internally determined, can only correct errors inside it. This may be a cause of filter divergence and give a rationale for the need to artificially increase the rank of the error covariance in EnKFs (e.g. by additive inflation).

In 4DVar-AUS, the dimension of the reduced subspace is kept fixed to an arbitrary number and varying this number can help deal with errors that are not exactly confined in the unstable and neutral subspace. Indeed, when nonlinearity becomes important (e.g. for the larger observation errors and longer windows in the experiments of [Trevisan *et al.*, 2010]) and errors develop in a larger subspace, it is useful to use a larger value for the fixed reduced subspace dimension. The above reasoning can also explain the superiority of 4DVar with respect to Kalman Filters in the presence of important nonlinearities.

An important aspect regarding the data assimilation algorithms for chaotic systems is that the knowledge of the Lyapunov vectors permits a wise use of adaptive or targeted observations. Indeed for each dynamical condition it is apparent how the most useful measurements are those located in the spatial locations where the unstable structures have larger components. Both EKF-AUS and 4DVar-AUS algorithms lead as a by-product a basis of the unstable and neutral manifolds of the system. This means that at any forecast time one can look for the locations where the most unstable perturbation vectors describing the error have larger components. This approach may be fruitful in operative NWP where adaptive observations based on the indication of the breeding method have already produced good results [Bishop & Toth, 1999].

There are at least two sectors where the results obtained so far can be fruitful both for theoretical and applied development. The first is the construction of Ensemble Forecasting Systems. We have already discussed how initial conditions, the outcome of the combination of forecasts and observations to form the analysis, depend on how the assimilation scheme deals with forecast errors in the different, stable and unstable directions. One can view Data assimilation and Ensemble Forecasting as a unified system where the number and the structure of perturbations of the forecast ensemble are an outcome of the assimilation process, while the forecast perturbations are an outcome of the assimilation. The AUS approach to data assimilation operates on a choice of the dimension and subspace for the assimilation both in the 4DVar and EKF context so that the number and the structure of perturbations can be a built-in part of the algorithm in a unified assimilation-forecast cycle approach.

The second is data assimilation in the presence of multiple-scale instabilities. Assimilation and ensemble prediction schemes, able to cope with convective systems, are currently being developed for nonhydrostatic limited area models. However, given the rate of increase of computer power, global NWP models that simultaneously predict all, from convective to planetary, scales may soon become a reality. The existing data assimilation algorithms are not adequate to simultaneously control the growth

of errors due to small scale fast growing and large scale slower unstable modes. For instance, tuning the length of the time window of 4DVar on convective motions, would penalize its efficiency with respect to synoptic scale flow. We think that, when constructing assimilation schemes for systems with multiple scale instabilities, progress can be expected if conscious choices, based on the knowledge of the chaotic properties, are made.

References

Bannister, R. N. [2008] "A review of forecast error covariance statistics in atmospheric variational data assimilation. I: Characteristics and measurements of forecast error covariances," *Q. J. Roy. Met. Soc.* **134**, 1951–1970.

Benettin, G., Galgani, L., Giorgilli, A. & Strelcyn, J. M. [1980] "Lyapunov characteristic exponents for smooth dynamical systems and for Hamiltonian systems: A method for computing them," *Meccanica* **15**, 9–21.

Berner, J., Shutts, G., Leutbecher, M. & Palmer, T. N. [2009] "A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system," *J. Atmos. Sci.* **66**, 603–626.

Bishop, C. H. & Toth, Z. [1999] "Ensemble transformation and adaptive observations," *J. Atmos. Sci.* **56**, 1748–1765.

Blum, J., Le Dimet, F.-X. & Navon, I. [2008] "Data assimilation for geophysical fluids, handbook of numerical analysis," Special Volume on *Computational Methods for the Ocean and the Atmosphere*, eds. Temam, R. & Tribbia, J. (Elsevier, NY), pp. 1–63.

Bougeault, P., Toth, Z., Bishop, C., Brown, B., Burridge, D., Chen, D.-H., Ebert, B., Fuentes, M., Hamill, T. M., Mylne, K., Nicolau, J., Paccagnella, T., Park, Y.-Y., Parsons, D., Raoult, B., Schuster, D., Silva Dias, P., Swinbank, R., Takeuchi, Y., Tennant, W., Wilson, L. & Worley, S. [2010] "The THORPEX interactive grand global ensemble (TIGGE)," *Bull. Amer. Meteor. Soc.* **191**, 59–172.

Brown, R., Bryant, P. & Abarbanel, D. I. [1991] "Computing the Lyapunov spectrum of a dynamical system from an observed time series," *Phys. Rev. A* **43**, 2787.

Buizza, R., Miller, M. & Palmer, T. N. [1999] "Stochastic representation of model uncertainties in the ECMWF ensemble prediction system," *Q. J. R. Meteorol. Soc.* **125B**, 2887–2908.

Cai, M., Kalnay, E. & Toth, Z. [2003] "Bred vectors of the Zebiak–Cane model and their potential application to ENSO predictions," *J. Climate* **16**, 40–56.

Candille, G. & Talagrand, O. [2005] "Evaluation of probabilistic prediction systems for a scalar variable," *Q. J. R. Meteorol. Soc.* **131**/609, 2131–2150.

Carrassi, A., Trevisan, A., Descamps, L., Talagrand, O. & Uboldi, F. [2008a] "Controlling instabilities along a 3DVar analysis cycle by assimilating in the unstable subspace: A comparison with the EnKF," *Nonlin. Proc. Geophys.* **15**, 503–521.

Carrassi, A., Vannitsem, S. & Nicolis, C. [2008b] "Model error and sequential data assimilation: A deterministic formulation," *Quart. J. Roy. Meteor. Soc.* **134**, 1297–1313.

Carrassi, A. & Vannitsem, S. [2010] "Accounting for model error in variational data assimilation: A deterministic formulation," *Mon. Wea. Rev* **138**, 3369–3386.

Daley, R. [1991] *Atmospheric Data Analysis* (Cambridge University Press).

Daley, R. [1992] "The effect of serially correlated observation and model error on atmospheric data assimilation," *Mon. Wea. Rev.* **120**, 164–177.

Danforth, C., Kalnay, E. & Miyoshi, T. [2007] "Estimating and correcting global weather model error," *Mon. Wea. Rev.* **135**, 281–299.

DelSole, T. & Hou, A. Y. [1999] "Empirical correction of a dynamical model. Part I: Fundamental issues," *Mon. Wea. Rev.* **127**, 2533–2545.

Eckmann, J. P. & Ruelle, D. [1985] "Ergodic-theory of chaos and strange attractors," *Rev. Mod. Phys.* **57**/3, 617–656.

Evensen, G. [1994] "Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics," *J. Geophys. Res.* **99**, 10143–10162.

Evensen, G. [2003] "The ensemble Kalman filter: Theoretical formulation and practical implementation," *Ocean Dynamics* **53**, 343–367.

Friedland, B. [1969] "Treatment of bias with recursive filtering," *IEEE Trans. Autom. Contr.* **14**, 359–367.

Fukumori, I. [2002] "A partitioned Kalman filter and smoother," *Mon. Wea. Rev.* **130**, 1370–1383.

Ghil, M. & Malanotte Rizzoli, P. [1991] *Data Assimilation in Meteorology and Oceanography*, ed. Saltzmann, B. (Academic Press), pp. 141–266.

Gustafsson, N. [2007] "Discussion on '4DVar or EnKF'?" *Tellus A* **59**, 774–777.

Hamill, T. M. & Whitaker, J. S. [2011] "What constrains spread growth in forecasts initialized from ensemble Kalman filters?" *Mon. Wea. Rev.* **139**, 117–131.

Harlim, J., Oczkowski, M., Yorke, J. A., Kalnay, E. & Hunt, B. [2005] "Convex error growth patterns in a global weather model," *Phys. Rev. Lett.* **94**, 228501.

Ide, K., Courtier, P., Ghil, M. & Lorenc, A. [1997] "Unified notation for data assimilation: Operational, sequential and variational," *J. Meteorol. Soc. Jpn.* **75**, 181–189.

Jazwinski, A. H. [1970] *Stochastic Processes and Filtering Theory* (Academic Press, NY).

Kalman, R. [1960] "A new approach to linear filtering and prediction problems," *Trans. ASME J. Basic Eng.* **82**, 3545.

Kalnay, E. & Dalcher, A. [1987] "Forecasting forecast skill," *Mon. Wea. Rev.* **115**, 349–356.

Kalnay, E. [2003] *Atmospheric Modeling, Data Assimilation and Predictability* (Cambridge University Press, Cambridge, UK).

Kalnay, E., Li, H., Miyoshi, T., Yang, S.-C. & Ballabrera-Poy, J. [2007] "4DVar or ensemble Kalman filter?" *Tellus A* **59**, 758–773.

Lacarra, J. G. & Talagrand, O. [1988] "Short-range evolution of small perturbations in a barotropic model," *Tellus A* **40**, 81–95.

Le Dimet, F. X. & Talagrand, O. [1986] "Variational algorithms for analysis and assimilation of meteorological observations — Theoretical aspects," *Tellus* **38**/2, 97–110.

Legras, B. & Vautard, R. [1996] "A guide to Lyapunov vectors," in *Predictability*, Vol. I, ed. Palmer, T., ECMWF Seminar, ECMWF (Reading, UK), pp. 135–146.

Leith, C. E. [1971] "Atmospheric predictability and two-dimensional turbulence," *J. Atmos. Sci.* **28**, 145–161.

Leith, C. E. & Kraichnan, R. H. [1972] "Predictability of turbulent flows," *J. Atmos. Sci.* **29**, 1041–1058.

Leith, C. E. [1978] "Objective methods for weather prediction," *Ann. Rev. Fluid Mech.* **10**, 107–128.

Leutbecher, M. & Palmer, T. N. [2008] "Ensemble forecasting," *J. Comput. Phys.* **227**, 3515–3539.

Lilly, D. K. [1990] "Numerical prediction of thunderstorms has its time come?" *Q. J. Roy. Met. Soc.* **116**, 779–798.

Li, H., Kalnay, E., Miyoshi, T. & Danforth, C. M. [2009] "Accounting for model errors in ensemble data assimilation," *Mon. Wea. Rev.* **137**, 3407–3419.

Lorenc, A. C. [2003] "The potential of the ensemble Kalman filter for NWP — A comparison with 4DVAR," *Q. J. R. Meteorol. Soc.* **129**, 3183–3293.

Lorenz, E. N. [1963] "Deterministic nonperiodic flow," *J. Atmos. Sci.* **20**, 130–141.

Lorenz, E. N. [1965] "A study of the predictability of a 28-variable atmospheric model," *Tellus* **17**, 321–333.

Lorenz, E. N. [1969] "Atmospheric predictability as revealed by naturally occurring analogues," *J. Atmos. Sci.* **26**, 636–646.

Lorenz, E. N. [1982] "Atmospheric predictability experiments with a large numerical model," *Tellus* **34**, 505–513.

Lorenz, E. N. [1984] "The local structure of a chaotic attractor in four dimensions," *Physica D* **13**, 90–104.

Lorenz, E. N. [1986] "Atmospheric models as dynamical systems," in *Perspectives in Nonlinear Dynamics* (World Scientific, Singapore), pp. 1–17.

Lorenz, E. N. [1993] *The Essence of Chaos* (University of Washington Press, Washington).

Lorenz, E. N. [1996] "Predictability: A problem partly solved," *Proc. Seminar Predictability. European Center for Medium-Range Weather Forecasting* (Shinfield Park, Reading, UK), pp. 1–18.

Lorenz, E. N. [2005] "A look at some details of the growth of initial uncertainties," *Tellus A* **57**, 1–11.

Miller, R. N., Ghil, M. & Gauthier, F. [1994] "Advanced data assimilation in strongly nonlinear dynamical systems," *J. Atmos. Sci.* **51**, 1037–1056.

Mitchell, H. L. & Houtekamer, P. L. [2000] "An adaptive ensemble Kalman filter," *Mon. Wea. Rev.* **128**, 416–433.

Mukougawa, I., Kimoto, M. & Yoden, S. [1991] "A relationship between local error growth and quasi-stationary states: A case study in the Lorenz system," *J. Atmos. Sci.* **48**, 1231–1237.

Nicolis, C. [1992] "Probabilistic aspects of error growth in atmospheric dynamics," *Quart. J. Roy. Meteor. Soc.* **118**, 553–568.

Nicolis, C. [2003] "Dynamics of model error: Some generic features," *J. Atmos. Sci.* **60**, 2208–2218.

Nicolis, C. [2004] "Dynamics of model error: The role of unresolved scales revisited," *J. Atmos. Sci.* **61**, 1740–1753.

Nicolis, C. [2005] "Can error source terms in forecasting models be represented as Gaussian Markov noises?" *Quart. J. Roy. Meteorol. Soc.* **131**, 2151–2170.

Nicolis, C. & Nicolis, G. [1993] "Finite time behavior of small errors in deterministic chaos and Lyapunov exponents," *Int. J. Bifurcation and Chaos* **3**, 1339–1342.

Nicolis, C., Perdigao, R. & Vannitsem, S. [2009] "Dynamics of prediction errors under the combined effect of initial condition and model errors," *J. Atmos. Sci.* **66**, 766–778.

Oseledec, V. I. [1968] "A multiplicative ergodic theorem, Lyapunov characteristic numbers for dynamical systems," *Trans. Moscow Math. Soc.* **19**, 197–231.

Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A. V., Kostelich, E. J., Corazza, M., Kalnay, E., Patil, D. J. & Yorke, J. A. [1995] "A local ensemble Kalman filter for atmospheric data assimilation," *Tellus A* **56**/5, 415–428.

Palmer, T. N., Gelaro, R., Barkmeijer, J. & Buizza, R. [1998] "Singular vectors, metrics, and adaptive observations," *J. Atmos. Sci.* **55**, 633–653.

Pham, D. T., Verron, J. & Roubaud, M. C. [1998] "A singular evolutive extended Kalman filter for data assimilation in oceanography," *J. Marine Syst.* **16**, 323–340.

Pires, C., Vautard, R. & Talagrand, O. [1996] "On extending the limits of variational assimilation in nonlinear chaotic systems," *Tellus A* **48**, 96–121.

Sasaki, Y. [1970] "Some basic formalism in numerical variational analysis," *Mon. Wea. Rev.* **98**, 875–883.

Schubert, S. D. & Suarez, M. [1989] "Dynamical predictability in a simple general circulation model: Average error growth," *J. Atm. Sci.* **46**/3, 353–370.

Smagorinsky, J. [1963] "General circulation experiments with the primitive equations," *Mon. Wea. Rev.* **91**, 99–164.

So, P., Ott, E. & Dayawansa, W. P. [1994] "Observing chaos: Deducing and tracking the state of a chaotic system from limited observation," *Phys. Rev. E* **49**, 2650–2660.

Swanson, K. L., Palmer, T. N. & Vautard, R. [2000] "Observational error structures and the value of advanced assimilation techniques," *J. Atmos. Sci.* **57**, 1327–1340.

Szunyogh, I., Kalnay, E. & Toth, Z. [1997] "A comparison of Lyapunov vectors and optimal vectors in a low resolution GCM," *Tellus A* **49**, 200–227.

Talagrand, O. [1997] "Assimilation of observations, an introduction," *J. Met. Soc. Jap.* **75**/1B, 191–209.

Talagrand, O. & Courtier, P. [1987] "Variational assimilation of observations with the adjoint vorticity equations," *Q. J. Roy. Met. Soc.* **113**, 1311–1328.

Thornton, C. L. & Bierman, G. J. [1976] "A numerical comparison of discrete Kalman filtering algorithms: An orbit determination case study," *JPL Technical Memorandum* **33**, Pasadena, p. 771.

Tippet, M. K., Cohn, S. E., Todling, R. & Marchesin, D. [2000] "Conditioning of the stable, discrete-time Lyapunov operator," *SIAM J. Matrix Anal. Appl.* **22**, 56–65.

Todling, R. & Cohn, S. E. [1994] "Suboptimal schemes for atmospheric data assimilation based on the Kalman filter," *Mon. Weather Rev.* **122**, 2530–2557.

Toth, Z. & Kalnay, E. [1993] "Ensemble forecasting at NMC: The generation of perturbations," *Bull. Am. Met. Soc.* **74**/12, 2317–2330.

Toth, Z. & Kalnay, E. [1997] "Ensemble forecasting at NCEP and the breeding method," *Mon. Wea. Rev.* **125**/12, 3297–3319.

Toth, Z. & Peña, M. [2007] "Data assimilation and numerical forecasting with imperfect models: The mapping paradigm," *Physica D* **230**, 146–158.

Tremolet, Y. [2007] "Model-error estimation in 4D-Var," *Q. J. Roy. Meteorol. Soc.* **133**, 1267–1280.

Trevisan, A., Malaguzzi, T. & Fantini, M. [1992] "On Lorenz's law for the growth of large and small errors in the atmosphere," *J. Atmos. Sci.* **49**, 713–719.

Trevisan, A. [1993] "Impact of transient error growth on global average predictability measures," *J. Atmos. Sci.* **50**, 1016–1028.

Trevisan, A. & Legnani, R. [1995] "Transient error growth and local predictability: A study in the Lorenz system," *Tellus A* **47**, 103–117.

Trevisan, A. & Pancotti, F. [1998] "Periodic orbits, Lyapunov vectors, and singular vectors in the Lorenz system," *J. Atmos. Sci.* **55**, 390–398.

Trevisan, A. & Uboldi, F. [2004] "Assimilation of standard and targeted observations within the unstable subspace of the observation-analysis-forecast cycle system," *J. Atmos. Sci.* **61**, 103–113.

Trevisan, A., D'Isidoro, M. & Talagrand, O. [2010] "Four-dimensional variational assimilation in the unstable subspace and the optimal subspace dimension," *Quar. J. Roy. Met. Soc.* **136**, 487–496.

Trevisan, A. & Palatella, L. [2010] "The extended Kalman filter and its reduction to the unstable subspace," Poster contribution presented at *Ecodyc2010*, Dresden (Germany). http://www.mpipks-dresden.mpg.de/~ecodyc10/Contributions/Trevisan.pdf.

Trevisan, A. & Palatella, L. [2011] "On the Kalman filter error covariance collapse into the unstable subspace," *Nonlin. Proc. Geophys.* **18**, 243–250.

Tribbia, J. J. & Baumhefner, D. P. [2004] "Scale interactions and atmospheric predictability: An updated perspective," *Mon. Wea. Rev.* **132**, 703–713.

Tsuyuki, T. & Miyoshi, T. [2007] "Recent progress of data assimilation methods in meteorology," *J. Met. Soc. Japan B* **85**, 331–361.

Uboldi, F. & Trevisan, A. [2006] "Detecting unstable structures and controlling error growth by assimilation of standard and adaptive observations in a primitive equation ocean model," *Nonlin. Proc. Geophys.* **16**, 67–81.

Vannitsem, S. & Nicolis, C. [1997] "Lyapunov vector and error growth patterns in a T21L3 quasigeostrophic model," *J. Atmos. Sci.* **54**, 347–361.

Vidard, P. A., Piacentini, A. & Le Dimet, F.-X. [2004] "Variational data analysis with control of the forecast bias," *Tellus* **56**, 177–188.

Wei, M., Toth, Z., Wobus, R. & Zhu, Y. [2008] "Initial perturbations based on the ensemble transform (ET) technique in the NCEP global operational forecast system," *Tellus A* **60**, 62–79.

Whitaker, J. S. & Loughe, A. F. [1998] "The relationship between ensemble spread and ensemble mean skill," *Mon. Wea. Rev.* **126**, 3292–3302.

Wolfe, C. L. & Samelson, R. M. [2007] "An efficient method for recovering Lyapunov vectors from singular vectors," *Tellus A* **59**, 355–366.